Transcriptional regulation of the *Ccn2* gene

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

By

Stephanie Lauren Frost

February 2019

Acknowledgements

Completing this PhD would not have been possible without the help of many people and I hope that my work has done them proud. I would firstly like to thank the University of Liverpool for funding my project.

I would like to express my profound gratitude to my primary supervisor Professor George Bou-Gharios, for giving me the project to start with, and although it has been challenging at times you have always pushed me to be a better scientist. My ambition to dedicate my career to transcriptional regulation is a testament to the inspiration that you have given me in completing this thesis. I would also like to thank my second supervisor Professor Eithne Comerford and third supervisor Dr Blandine Poulet for their guidance throughout the project. I would also like to thank Dr Ke Liu for all of his help in creating the transgenics, and Professor David Young and Dr Louise Reynard from Newcastle University for teaching me CRISPR and EMSA respectively. I am grateful for the help that I received from Dr Ian Li, Dr Amy Horwell and Dr Craig Keenan in my laboratory work, with a few laughs along the way. I would also like to thank Gemma Charlesworth and Mandie Prior for trying to teach me the dark art of histology in the quest for a pretty section image.

Anybody who knows me can appreciate how bad I am with feelings and sentimentality so here goes...I would like to thank Dr Alan Carter for random conversations about Radio 4 documentaries, hysterical jokes, stories and Fat Fridays, and also (the soon to be Dr) Michele Scotto Di Mase for letting me corrupt him with terrible English phrases and pints, Pal. I will forever be indebted to (the soon to be Dr) Francesca Johnson Brito, the past couple of years living with you have been some of the most enjoyable of my life and you are the best friend that anybody could possibly wish for. Thank you for all of your support, our escapades and adventures, and more importantly for putting up with me!

I am wholeheartedly grateful for the unconditional love and encouragement that my mum Alison, Brother Sam and Sister Josie have given me over the past three years. Words cannot justify how much you mean to me and have inspired me (with a hint of sibling rivalry), and I would like to dedicate this work to them. I would additionally like to thank my grandparents Barry and Valerie and stepfather Tony for all of their support.

If you find the bad pun, that’s the true accomplishment in this work.

Table of contents

Transcriptional regulation of the *Ccn2* gene

Stephanie Frost

Abstract

Cellular Communication Network Factor 2 (CCN2) is a matricellular protein which functions in many tissues, and is most notably expressed by chondrocytes, with knockout of the *Ccn2* gene expression causing severe chondrodysplasia in mice. Regulation at the prerequisite stage of transcription by non-coding genomic elements is fundamental in the expression of every gene. Until now, the capacity for *cis*-acting regulatory modules to control the expression of *Ccn2* has been obscure. The current project sought to delineate the capacity for *cis*-acting enhancer regions to regulate *Ccn2* transcription within highly specific temporospatial contexts.

A 300 kilobase intergenic genomic region upstream of *Ccn2* was examined *in silico* in order to identify putative enhancer regions based on chromatin characteristics of enhancers; histone modification, DNase I hypersensitivity and interspecies evolutionary conservation of DNA sequence. Transgenic mice were created using constructs consisting of each putative enhancer region driving the expression of a *LacZ* reporter gene in conjunction with a silent Hsp68 minimal promoter. Expression of the protein product of *LacZ*, β-galactosidase was assayed at several developmental time-points in order to determine whether candidate enhancers were able to control gene transcription. Five enhancer regions of murine *Ccn2* expression, located -230kb, -198kb, -148kb, -137kb and -102kb upstream of the gene each drove *LacZ* expression in a tissue-specific manner at embryonic day E15.5. -230kb drove transgene activity within osteoblasts, whereas -198kb, -148kb and -137kb all exhibited function within chondrocytes. The -102kb enhancer was active within the superficial vasculature. More comprehensive examination of the -148kb enhancer revealed function in adulthood, and that a truncated region of this sequence is also capable of enhancing gene transcriptional output. Enhancer sequences were also examined *in vitro* using Electrophoretic Mobility Shift Assay in order to test the capacity of transcription factors of interest to bind to enhancers and modulate function. Master regulator of chondrocyte physiology, the transcription factor SOX9 was found to bind to sequences within the -137kb, -148kb and -230kb enhancers.

The findings described herein are the first to characterise the role of *cis*-acting enhancer regions in the transcription of *Ccn2* within the murine genome. Understanding the mechanisms that underpin temporospatial control of *Ccn2* expression will inform both the characterisation and amelioration of profound pathological conditions that result from loss of *CCN2* transcriptional regulation such as osteoarthritis, fibrosis and cancer in humans.

Index of Figures

Index of Tables

List of abbreviations

µg	Microgram	**CTGF**	Connective tissue growth factor
µL	Microlitre	**CTP**	Cytidine triphosphate
µm	Micrometre	**dATP**	deoxyadenosine triphosphate
µM	Micromolar	**dCTP**	deoxycytidine triphosphate
µmol	micromole	**DDR2**	Discoidin domain receptor 2
ADAMTS	and a disintegrin and metalloproteinase with a thrombospondin motif	**DEPC**	Diethyl pyrocarbonate
AFAP1	actin filament-associated protein	**dGTP**	deoxyguanosine triphosphate
ALP	Alkaline phosphatase	**DMEM**	Dulbecco's Modified Eagle Medium
AP-1	Activator protein-1	**DNA**	Deoxyribonucleic acid
ASPA	Animals (Scientific Procedures) Act	**DNase I**	Deoxyribonuclease I
ATP	Adenosine triphosphate	**DNMT1**	DNA methyltransferase 1
att	Gateway technology recombination site	**dNTP**	Deoxyribonucletotide triphosphates
BAC	Bacterial artificial chromosome	**dpc**	Days post-coitum
BCE	Basal control element	**DPE**	downstream promoter element
BED	Browser extensible data	**DSB**	Double strand break
BGLAP	Bone gamma-carboxyglutamate protein	**dTTP**	deoxythymidine triphosphate
BLAST	Basic Local Alignment Search Tool	**E8.5**	Embryonic day 8.5
BMP	Bone morphogenic proteins	**E9**	Embryonic day 9
bp	Base pair	**E11.5**	Embryonic day 11.5
BRE	TFIIB recognition element	**E12.5**	Embryonic day 12.5
CAESAR	*cis*-acting element of structure-anchored repression	**E13.5**	Embryonic day 13.5
CBP	cAMP response element-binding protein binding protein	**E14.5**	Embryonic day 14.5
CCN2	Cellular communication network factor 2	**E15.5**	Embryonic day 15.5
ChIA-Pet	Chromatin Interaction by Paired-End Tag Sequencing	**E16.5**	Embryonic day 16.5
ChIP	Chromatin immunoprecipitation	**E17.5**	Embryonic day 17.5
ChIP-Seq	Chromatin immunoprecipitation sequencing	***E.coli***	*Escherichia coli*
Col11a2	Collagen type XI a2	**ECM**	Extracellular matrix
Col1a1	Collagen type I a1	**EDTA**	ethylenediaminetetraacetic acid
Col2a1	Collagen type 2 a1	**EMSA**	Electrophoretic mobility shift assay
Col9a2	Collagen type IX a2	**ENCODE**	Encyclopaedia of DNA Elements
COMP	Cartilage oligomeric matrix protein	**ERK**	Extracellular signal-regulated kinases
CR	Chromatin remodelling protein	**eRNA**	Enhancer RNA
CRISPR-Cas9	clustered regularly interspersed short palindromic repeats-Caspase 9	**ET-1**	Endothelin 1
CRM	*Cis*-regulatory module	**EtBr**	Ethidium bromide
CT	C-terminal	**F_0**	Founder
CTCF	CCCTC-binding factor	**F_1**	Progeny of founder

FBS	Foetal bovine serum	**Inr**	Initiator
Fli-1	friend leukaemia integration-1	**IP**	Intraperitoneal
fmol	Femtomole	**IVD**	Intervertebral disc
FSH	Follicle stimulating hormone	**JNK**	c-Jun N-terminal kinase
g	Gravitational force	**kb**	kilobase
g	Gram	**KCl**	Potassium chloride
GC	guanine-cytosine	**KOH**	Potassium hydroxide
GDF	Growth differentiation factors	**L**	Litre
GEO	Gene Expression Omnibus	***LacZ***	β-galactosidase gene
GFP	Green fluorescent protein	**LARII**	Luciferase assay reagent II
GH	Growth hormone	**LB**	Lennox L
gRNA	Guide RNA	**LH**	Luteinising hormone
GTF	General transcription factor	**LICR**	Ludwig Institute for Cancer Research
GTP	Guanosine triphosphate	**lincRNA**	Long intergenic non-coding RNA
GW	Gateway	**LRP1**	Low-density lipoprotein receptor-related protein 1
H2	Histone 2	***Luc2***	Firefly luciferase gene
H3	Histone 3	**M**	Molar
H3K122ac	acetylation of lysine 122 of histone 3	**MAPK**	Mitogen-activated protein kinase
H3K27ac	Acetylation of lysine 27 residue of histone 3	**mb**	megabase
H3K27me3	trimethylation of lysine 27 residue of histone 3	**MC3T3**	Immortalised mouse calvaria cell line
H3K4me1	Monomethylation of lysine 4 residue of histone 3	**MCS**	Multiple cloning site
H3K4me3	Trimethylation of lysine 4 residue of histone 3	**MEFs**	Mouse embryonic fibroblasts
H3K64ac	acetylation of lysine 64 of histone 3	**MEK**	Mitogen-associated protein kinase kinase
H3K79me3	trimethylation of lysine 79 of histone 3	**Mg**	Milligram
H4	Histone 4	**$MgCl_2$**	Magnesium chloride
HCl	Hydrochloride	**mL**	Millilitre
HCS	Human chondrosarcoma	**MLL3/4**	Myeloid/Lymphoid or Mixed-Lineage Leukaemia 3/4
HCS-24	hypertrophic chondrocyte-specific gene product-24	**MMP**	Matrix metalloproteinase
HEK293	Human embryonic kidney cell line	***Moxd1***	monooxygenase DBH-like 1
HIF	Hypoxia inducible factor	**mRNA**	Messenger RNA
HOX	Homeobox	**NaOH**	Sodium hydroxide
HPLC	High performance liquid chromatography	**NCBI**	National Center for Biotechnology Information
Hsp68	Heat shock protein 68	**NDR**	Nucleosome depleted region
HTB-94	Immortalised human chondrosarcoma cell line	**NELF**	Negative elongation factor
IGF	Insulin-like growth factor	**ng**	Nanogram
IGF-B	insulin-like growth factor protein binding	**NHEJ**	Non-homologous end joining
IGFBP-8	Insulin-Like Growth Factor-Binding Protein 8	**NIH3T3**	Immortalised mouse fibroblast cell line
Ihh	Indian hedgehog	**NLS**	Nuclear localisation signal

nM	Nanomolar
nm	Nanometre
nt	nucleotide
OA	osteoarthritis
Osx	Osterix
P7	Postnatal day 7
PAM	Protospacer adjacent motif
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pg	Picogram
PIC	Pre-initiation complex
pmol	Picomole
Pol II	RNA polymerase II
PRC2	Polycomb repressive complex 2
pSLF01	pGL4.10 firefly luciferase reporter plasmid containing *CCN2* promoter region
P-TEFb	Positive elongation factor b
PTHrP	Parathyroid hormone-related peptide
rCCN2	Recombinant Cellular communication network factor 2
Rluc	*Renilla* firefly gene
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	Revolutions per minute
Runx	Runt-related transcription factor
SDS	Sodium dodecyl sulfate
sgRNA	Single guide RNA
SOC	super optimal broth with catabolite repression
Sox	Sex determining Region Y- Box
Sp1	Specificity factor 1
SpCas9	*Streptococcus pyrogenes* Caspase 9
SPF	Specified pathogen free
SRC	proto-oncogene tyrosine-protein kinase SRC
TAD	Topologically associated domain
TAE	Tris-acetic acid-ethylenediaminetetraacetic acid
TBE	Tris-boric acid-ethylenediaminetetraacetic acid
TBP	TATA binding motif
Tβ-RE	TGF-β response element
TC28-i2	Immortalised human rib derived chondrocyte cell line
TCF-LEF	T-cell factor- lymphoid enhancer binding factor
TE	Tris-EDTA
TEMED	N, N, N', N'-Tetramethylethylenediamine
TF	Transcription factor
TFBS	Transcription factor binding site
TFII	Transcription factor II
TGF-β	Transforming growth factor β
TIMP	Tissue inhibitor of metalloproteinase
T_m	Melting temperature
TNFα	Tumour necrosis factor α
tracrRNA	*Trans*-activating CRISPR RNA
TRAP	Transcription Factor Affinity Prediction
TRENDIC	Transcription enhancer dominant in chondrocytes
TSP	Thrombospondin
TSS	Transcription start site
uaRNA	Upstream antisense RNA
UCSC	University of California Santa Cruz
UTP	Uridine triphosphate
UTR	Untranslated region
VEGF	Vascular endothelial growth factor
v/v	Volume by volume
vWC	von Willebrand factor type C
w/v	Weight by volume
WT	Wild-type
X-gal	5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside
YY1	Yin yang 1

1. Introduction

1.1 Skeletal development and articulation is dependent on cartilage

1.1.1 Cartilage and the chondrocyte

Cartilage is a connective tissue which functions in the establishment and articulation of the skeleton, it is avascular and aneural, with chondrocytes as the only type of cell resident in the tissue. Cartilage is formed from an extracellular matrix (ECM) secreted by chondrocytes that is enriched for polysaccharides, proteoglycans and fibrous proteins, such as aggrecan, hyaluronan, cartilage oligomeric matrix protein (COMP) and cellular communication network factor 2 (CCN2), in addition to collagen types II, IX, X, XI, VI, XII, and XIV (Heinegård 2009; Luo *et al.* 2017). The proportion of these components confers the properties of the three classes of cartilage; hyaline, fibrocartilage and elastic cartilage. Hyaline cartilage is marked by abundance of collagen type II and can be found in a temporary form in the anlage of the vertebrate skeleton, and permanently in the trachea and articular cartilage of synovial joints. This is in contrast to the fibrocartilage found in tendon and ligament attachment sites in addition to intervertebral discs (IVD), which contains a greater proportion of collagen type I lending this tissue greater rigidity. Finally, elastic cartilage contains elastic fibres that enable flexibility of the pinna and epiglottis (Hall 2015). In each type of cartilage chondrocyte behaviour must be tightly controlled in order to ensure physiological development and homeostasis.

Chondrocytes are of the mesenchymal cell lineage, and during embryonic development develop through the condensation of somitic mesoderm and lateral plate mesoderm progenitor cells (Maes and Kronenberg 2016). As with every other cell and tissue type, stringent regulation of gene expression underlies cartilage and chondrocyte identity and function. Transcription factor (TF) proteins are integral in this; controlling target gene expression through interaction with genomic regulatory sequences. SOX9 (Sex Determining Region Y- Box 9) has a major role in chondrogenesis and is viewed as the ‘master’ regulator of chondrocyte behaviour and function (Karsenty 2008). This transcription factor regulates the expression of many genes that are vital in chondrocyte behaviour and cartilage homeostasis. Chondrocytes at the terminal stage of differentiation- hypertrophy, are the only sub-population of chondrocyte to not express *SOX9*, with absence of messenger

ribonucleic acid (mRNA) in this cell type (Akiyama *et al.* 2002). SOX9 target genes include ECM components such as collagen types II (*Col2a1*), IX, X and collagen type XI (Bar Oz *et al.* 2016; Lefebvre and Dvir-Ginzberg 2017). Bernard *et al*. (2003) first described the capacity of SOX9 to dimerise in order to regulate the expression of cartilaginous collagens type IX (*Col9a2*) and type XI (*Col11a2*) through interaction with paired transcription factor binding sites (TFBS). Subsequently, many SOX9 target genes have been identified on the basis of TFBS within regulatory sequences (Oh *et al.* 2010). In addition, SOX9 is also capable of perpetuating further regulation of cartilage behaviour through control of genes encoding regulatory proteins; including *SOX9* itself and binding partners such as *SOX5* and *SOX6* (Akiyama *et al.* 2002; Yao *et al.* 2015). SOX5 and SOX6 synergise with SOX9, forming the 'chondrogenic Sox trio' and are required for chondrogenesis (Lefebvre and Smits 2005). The binding of SOX9 to target sequences is improved through interaction with SOX5/6, and recognition sites for the three proteins are clustered in genomic regulatory elements (Han and Lefebvre 2008; Liu and Lefebvre 2015). The profound importance of SOX9 in chondrocyte physiology is emphasised by the effect that loss of gene function has on cartilage. Bi *et al*. (2001) generated *Sox9* haploinsufficient mice which exhibited a severe cartilage related phenotype resulting in the bending of many skeletal structures and perinatal mortality. In humans, aberration of *SOX9* expression leads to chondrocyte and cartilage related pathologies including campomelic dysplasia and osteoarthritis (Yao *et al.* 2015).

In a seminal study, Wright *et al*. (1995) were the first to directly prove that SOX9 drives chondrogenesis, with *Sox9* mRNA observed within mouse embryos at embryonic day 9 (E9), coinciding mesenchymal condensation. As chondrogenic differentiation ensues in mice, early chondrocytes are distinguishable by embryonic day 11.5 (E11.5) (Lefebvre and Smits 2005). Akiyama *et al*. (2002) demonstrated that removal of *Sox9* expression prior to the induction of mesenchymal condensation prevents formation of cartilage and bone. In the same study, knockout of *Sox9* after mesenchymal cell condensation repressed chondrocyte proliferation and differentiation. SOX9 has been well established as promoting chondrocyte proliferation in addition to the secretion of key ECM components such as aggrecan and collagen type II, thereby establishing cartilaginous tissue identity (Chimal-Monroy *et al.* 2003). Dy *et al.* (2012) used conditional knockout of *Sox9* in pre-hypertrophic chondrocytes in the context of collagen type II expression to demonstrate that SOX9 reinforces lineage specificity in these cells, as removal of

Sox9 expression led to premature acquisition of osteoblastic phenotype. SOX9 is therefore capable of repressing the expression of genes non-specific to chondrocytes. This reinforces the findings of Oh *et al*. (2010) who found SOX9 binding sites in regulatory sequences for bone related genes osterix (*Osx*) and runt-related transcription factor 2 (*Runx2*), the expression of which are repressed in chondrocytes not undergoing differentiation. SOX9 is also required in cartilaginous tissue in the postnatal period, in order to ensure cartilage homeostasis. Henry *et al.* (2012) induced knockout of *Sox9* expression in an aggrecan expression linked and therefore cartilage specific manner in adulthood, observing aberration of articular, growth plate and intervertebral disc cartilage.

Sox9 function can be modulated through signalling of the Transforming Growth Factor β (TGF-β) superfamily of proteins, which also regulate cartilage and skeletogenesis (Wang *et al.* 2014). For example, TGF-β signalling stabilises and phosphorylates SOX9 thereby enhancing its function (Coricor *et al.* 2016). The TGF-β superfamily contains factors such as TGF-β, bone morphogenic proteins (BMPs) growth differentiation factors (GDFs) and Smads. Over the past 30 years TGF-β has been well established as a regulator of chondrogenesis, with early studies demonstrating that application of exogenous TGF-β to chick limb mesenchyme led to enhanced chondrogenesis and cartilage deposition (Kulyk *et al.* 1989; Leonard *et al.* 1991). TGF-β initiates further cascades of cellular signalling such as mitogen-activated protein kinases (MAPK) and extracellular signal-regulated kinases (ERK), which regulate the expression of proteins such as N-cadherin and fibronectin involved in mesenchymal condensation and adhesion in early chondrogenesis, (Song *et al.* 2007; Tuli *et al.* 2003). TGF-β also reinforces chondrocyte phenotype through Smad2/3 signal transduction which prevents premature hypertrophy and therefore terminal differentiation whilst cartilaginous tissue is being established (van der Kraan *et al.* 2009). BMPs are also important regulators of skeletogenesis (Minina *et al.* 2001). For example, Yi *et al*. (2000) created BMP receptor I B knockout mice which exhibited skeletal defects, primarily in digit development. The impairment of BMP signalling led to reduced chondrocyte proliferation and chondrogenesis in the phalanges, demonstrating that whilst this signalling was not crucial in the establishment of the skeleton on a gross scale, BMP signalling through this receptor is required in order to refine skeletal structure.

1.1.2 Skeletogenesis depends on tightly regulated ossification processes

The skeleton is a highly dynamic structure which is stringently regulated in order to ensure proficient development and maintenance in every vertebrate organism. Skeletal tissue originates from primitive mesenchymal tissue that undergoes differentiation to form chondro-osseous tissues during embryonic development and beyond (Ono *et al.* 2014). Efficient skeletogenesis depends on the coordination of vastly complex cellular signalling and regulatory interactions that mediate patterning of the axial and appendicular skeletal structures in a highly specific temporospatial manner (Liu *et al.* 2017). All skeletal tissue is formed through the processes of endochondral ossification and intramembranous ossification (Maes and Kronenberg 2016). The use of murine models has been an imperative tool in the elucidation of the molecular processes that underpin skeletal biology. For example, the key steps in chondrogenesis and endochondral ossification typically occur between embryonic days 11.5 to 15.5 (E11.5-E15.5) during mouse development (Kaufman 2003). However, much is still unknown about the intricate mechanisms that facilitate dynamism in the skeleton during development and beyond.

1.1.2.1 Endochondral ossification

The appendicular skeleton and several components of the axial skeleton such as the ribs and vertebrae are formed through the process of endochondral ossification, whereby a cartilaginous anlage develops before being replaced with osseous tissue (Mackie *et al.* 2008). As aforementioned, chondrocytes arise from mesenchymal progenitor cells and form the cartilaginous template from which endochondral bone is created. In mice, this structure is well stratified by E15.5 on the basis of chondrocyte differentiation state, which further reinforces regulation of skeletal development (Long and Ornitz 2013). Primary endochondral ossification ensues as chondrocytes move along this growth plate differentiation gradient through five major zones; resting, proliferating, pre-hypertrophic, hypertrophic and ossification centre (as illustrated in Figure 1.1) (Mackie *et al.* 2008). This process is repeated in the postnatal period with secondary ossification that allows further interstitial growth of endochondral bones. *Sox5*, *Sox6* and *Sox9* are all highly expressed throughout the majority of endochondral tissue, with the exception of the hypertrophic zone (Chimal-Monroy *et al.* 2003). This expression profile reinforces chondrocyte function, therefore ensuring that sufficient cartilage template is created

for efficient endochondral bone growth during development. For example, loss of the expression of SOX9 target *Col2a1* prevents endochondral ossification within long bones and leads to growth plate disorganisation (Li *et al.* 1995). This highlights the importance of efficient cartilage matrix deposition in endochondral bone formation.

Resting and proliferative chondrocytes

The resting zone is located at the ends of the endochondral bone anlage, the future epiphysis, and contains chondrocytes with round morphology that secrete typical cartilage ECM components such as aggrecan and collagen type II (Karsenty 2008). This hyaline cartilage is maintained to form articular cartilage upon completion of differentiation (Goldring 2012). The proliferative zone is found next to the resting region. Chondrocytes proliferate rapidly within this area and are found in clusters with flattened morphology (Mackie *et al.* 2008). These cells secrete ECM rich in collagen type VI and matrilin 1 (Goldring 2012).

Pre-hypertrophic chondrocytes

Moving along the growth plate and with more advanced cellular differentiation state, pre-hypertrophic chondrocytes are located next to proliferative chondrocytes and distinguished by columnar organisation and attenuation of cellular proliferation (Liu *et al.* 2017). These cells are observed by E13.5 in mice (St-Jacques *et al.* 1999). Parathyroid hormone-related peptide (PTHrP) and Indian hedgehog (Ihh) signalling are critical regulators of pre-hypertrophic chondrocyte behaviour (Mackie *et al.* 2008). St-Jacques *et al*. (1999) created $Ihh^{-/-}$ mice, which exhibited abnormality in endochondral ossification. Mutants were similar to wild-type (WT) controls at E12.5; before the onset of pre-hypertrophic chondrocyte signalling, yet demonstrated significantly shorter limbs as endochondral ossification proceeded. Mutant chondrocytes were not organised in the pre-hypertrophic zone and chondrocyte proliferation was impaired. Mineral deposition had occurred in a similar fashion to WT controls, therefore disruption in hypertrophy was the cause of decreased bone length (St-Jacques *et al.* 1999). Ihh controls PTHrP signalling through the regulation of PTHrP expression (Vortkamp *et al.* 1996). PTHrP ligand is produced by chondrocytes of the proliferative zone and perichondrium, but signalling is transduced in pre-hypertrophic chondrocytes which express the PTHrP receptor (Long and Ornitz 2013; Mackie *et al.* 2008). PTHrP prevents chondrocyte

hypertrophy (Kobayashi *et al.* 2002). Huang *et al.* (2001) demonstrated that PTHrP signalling leads to phosphorylation of SOX9 in pre-hypertrophic chondrocytes. This promotes the retention of chondrocyte cellular phenotype and represses premature differentiation (Huang *et al.* 2001). This is exemplified by the observation that *PTHrP*$^{-/-}$ mice exhibit increased formation of endochondral bone, with larger zones of hypertrophic chondrocytes (Serra *et al.* 1999). PTHrP in turn limits Ihh prevalence through limiting the proportion of cells transitioning into hypertrophy, this negative feedback loop allows stringent regulation of chondrocyte differentiation (Mackie *et al.* 2008).

Chondrocyte hypertrophy and primary ossification

By embryonic day 15.5, pre-hypertrophic cells transition into terminal differentiation in the hypertrophic zone, which provides a front with the primary ossification centre where trabecular bone is formed (as illustrated in Figure 1.1) (Long and Ornitz 2013). Hypertrophic chondrocytes are distinguished with larger cellular volume and elongated morphology compared to pre-hypertrophic chondrocytes (Lefebvre and Smits 2005). This transformation of cellular phenotype is accompanied by a marked change in gene expression profile, the composition of the cartilage ECM secreted by the chondrocytes is modified due to repression of the expression of typical cartilaginous markers such as *Col2a1* and a change to collagen type X (*Col10a1*) production (Lefebvre and Smits 2005). The cartilage matrix becomes mineralised, or is resorbed as blood vessels, osteoclasts and osteoblasts invade from the periosteal bone collar (Mackie *et al.* 2008). Matrix remodelling occurs with invading osteoclasts degrading the cartilage matrix, which is subsequently replaced with trabecular bone matrix secreted by osteoblasts which is enriched for collagen type I (*Col1a1*), and mineralised through further deposition of hydroxyapatite (Mackie *et al.* 2008; Maes *et al.* 2010).

Loss of *Sox9* expression is a notable hallmark of hypertrophic chondrocytes. Hattori *et al.* (2010) generated mice harbouring knock-in of *Sox9* expression in the context of *Col10a1* thereby leading to aberrant *Sox9* expression within hypertrophic chondrocytes. This caused repression of bone marrow formation, reduced vascularisation of cartilage and inhibition of the commitment to terminal state; therefore preventing ossification. However, Dy *et al.* (2012) showed that SOX9 protein does prevail in wild-type hypertrophic chondrocytes, and is in fact necessary for the induction of hypertrophy. He *et al.* (2016) built on this with the discovery that

SOX9 acts in conjunction with activator protein-1 (AP-1) proteins to drive the expression of hypertrophy associated genes such as *Col10a1*, with repressive transcription factor function in the absence of AP-1. Vascular endothelial growth factor (*VEGF*) expression is repressed by SOX9 in immature chondrocytes, ensuring that expression is specific to hypertrophic chondrocytes (Hattori *et al.* 2010). This protein mediates growth plate angiogenesis and therefore the blood vessel invasion within the avascular cartilage that is required for osteogenesis (Mackie *et al.* 2008).

RUNX2 and RUNX3 are major regulators of hypertrophic chondrocyte and osteoblast cellular behaviour and function (Yoshida *et al.* 2004). Komori *et al.* (1997) used a *Runx2*$^{-/-}$ murine model to demonstrate that RUNX2 function is of paramount importance in ossification. These mice died shortly after birth and exhibited gross impairment of ossification. RUNX2 activity increases the expression of genes associated with chondrocyte terminal differentiation and osteoblastogenesis including *Col10a1*, *VEGF* and *matrix metalloproteinases-9* and *13* (*MMP9* and *MMP13*); thereby promoting the transition from cartilage anlage to bone (Lefebvre and Smits 2005; Ortega *et al.* 2004). MMP9 and MMP13 degrade cartilaginous ECM components such as collagen type II and aggrecan which enables blood vessel and osseous cell invasion (Mackie *et al.* 2008).

Secondary ossification

Secondary ossification occurs during postnatal development and allows further endochondral bone growth (Long and Ornitz 2013). Within the long bones, secondary ossification centres are found in the epiphyseal ends, as demonstrated in Figure 1.1. Chondrocytes within these regions undergo cell cycle arrest and hypertrophy, in a similar manner as aforementioned, with invasion of blood vessels and osteoblasts leading to the conversion of the cartilaginous anlage into osseous tissue (Wang *et al.* 2014). A band of chondrocytes is retained after the resolution of the primary and secondary ossification centres (Mackie *et al.* 2008). This metaphyseal growth plate provides a reservoir of chondrocytes that undergo hypertrophy in an ordered fashion; facilitating further interstitial growth of the bone until the cessation of skeletal development in adulthood (Kronenberg 2003). The processes of primary and secondary ossification are illustrated in Figure 1.1.

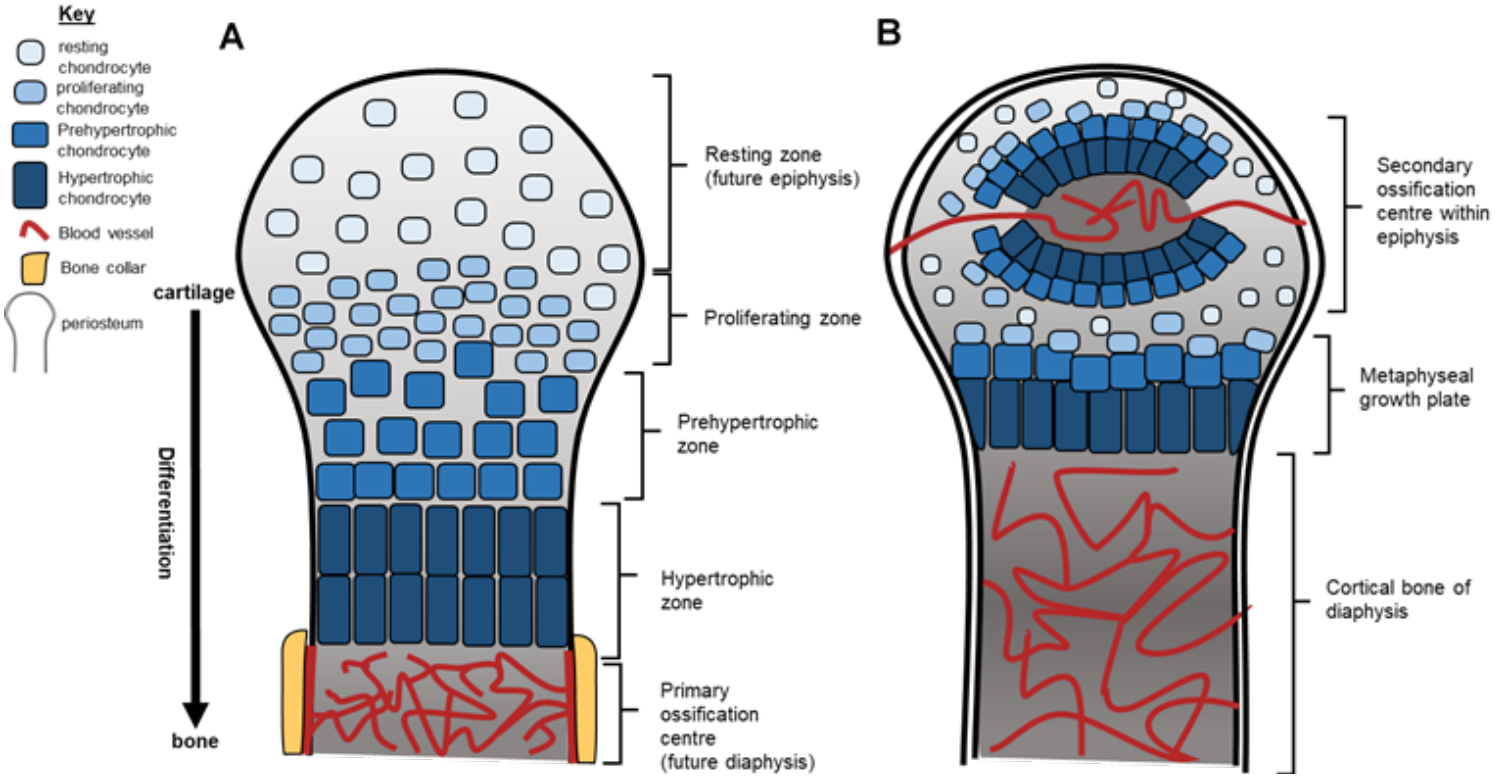


Figure 1.1: Schematic representation of endochondral ossification of a long bone. The growth plate is stratified on the basis of cellular differentiation state. During primary ossification (A), chondrocytes in the resting zone are in a primitive state of differentiation. This is also true of proliferative chondrocytes which have a flattened morphology and replicate rapidly. Proliferative chondrocytes become organised as they transition into the pre-hypertrophic zone, forming columns of cells that are subject to PTHrP-Ihh signalling axis, which ensures adequate preliminary interstitial growth of the future bone. Cellular proliferation ceases as pre-hypertrophic cells move into the hypertrophic zone, and lose expression of SOX9, accompanying a change in matrix composition with expression of Col10a1. Hypertrophic chondrocytes undergo apoptosis or transdifferentiate into osseous cells in the primary ossification centre. This region is surrounded by the bone collar, from where blood vessels and osteoblasts invade. This results in matrix remodelling and mineralisation in the diaphysis of the future bone. This process is repeated in the secondary ossification centre which forms in the future epiphysis of the bone (B). The metaphyseal growth plate is retained during postnatal interstitial growth.

Hypertrophic chondrocyte cell lineage fate

Hypertrophic chondrocytes have classically been viewed as undergoing apoptosis, making way for osteoblastic cells to populate the endochondral framework and form trabecular bone (Kronenberg 2003). *In vivo* lineage tracing methodology has been an important approach in attempts to understand the cellular differentiation processes that underpin endochondral ossification. A long held consensus view was that osteoblasts originate from the perichondrium (Long and Ornitz 2013). For example, Maes *et al.* (2010) labelled osteoblast cells based on regulation of reporter gene in the context of osteoblast markers *Col1a1* and *Osx*. Labelled cells were initially observed within the periosteum at E14.5, yet by E16.5 were found within the osseous tissue. However, there were discrete differences in cellular localisation based on the two contexts of reporter gene expression. *Osx* targeted reporter expression occurred predominantly within trabecular bone, whereas reporter gene expression in the context of *Col1a1* was mainly found in cortical bone. Findings from this study support the notion that the vast majority of primary ossification centre osteoblasts originate from the perichondrium. The rigidity in this process was rather controversially brought into question by Ono *et al.* (2014) who also used a lineage tracing approach in the examination of osteoblastogenesis during endochondral ossification. The Cre recombinases used in this study were expressed within cartilaginous (*Col2a1*) and osseous (*Col1a1*) contexts leading to the expression of reporter gene where these promoters were utilised initially, in addition to subsequent daughter cells. Reporter genes were concurrently expressed within pre-osteoblast cells within the perichondrium, with RUNX2 as a driver of osteoblastic differentiation. Therefore chondrocyte cells within the perichondrium were the progenitors of these osteoblastic cells (Ono *et al.* 2014).

Further *in vivo* studies have demonstrated that hypertrophic chondrocytes are able to transdifferentiate into osteoblasts. Both Zhou *et al.* (2014) and Yang *et al.* (2014) used lineage tracing methodology specific to hypertrophic chondrocytes. The Cre-recombinase transgene created in both of these studies was under the control of the *Col10a1* promoter. This approach demonstrated that hypertrophic cells are progenitors for osseous cells within the primary ossification centre. Reporter protein was perpetuated within osseous cells of the primary ossification centre; if hypertrophic chondrocytes underwent apoptosis with no transdifferentiation this expression would have been lost. The concept of plasticity in skeletal cell differentiation was also explored by Mizuhashi *et al.* (2018) who examined the

prospect of cell lineage switching in the growth plate in further detail. This study used PTHrP based lineage tracing to demonstrate that chondrocytes within the resting zone of the growth plate go on to form columnar chondrocytes within the growth plate, eventually undergoing hypertrophy and terminal differentiation. Moreover, these cells gave rise to osteoblasts and bone marrow stromal cells rather than undergoing apoptosis. These precursor cells expressed stem cell markers therefore lending credence to the notion that these cells are progenitors of skeletal lineage cells (Mizuhashi *et al.* 2018). These studies highlight the intricacies involved in chondro-osseous cell lineage specification, in addition to the need for further examination of the protein function and molecular mechanisms that govern endochondral ossification.

1.1.2.2 Intramembranous ossification

Intramembranous ossification facilitates formation of the majority of the cranial and facial bones in addition to the clavicle during embryonic development (Maes and Kronenberg 2016). This process is simpler than endochondral ossification as osteoblast cells develop directly from mesenchymal progenitors, with no cartilaginous intermediate (Long and Ornitz 2013). The mesenchymal precursor cells condense and give rise to osteoblast cells which secrete osteoid matrix that becomes mineralised (Wu *et al.* 2016). RUNX2 is a major regulator of intramembranous ossification, promoting osteoblast differentiation from mesenchymal progenitors (Maes *et al.* 2010). *Runx2*$^{-/-}$ mice display a lack of intramembranous skeletal elements (Komori *et al.* 1997). These mice also do not express the aforementioned bone marker Osx; another transcription factor which induces osteoblast differentiation (Karsenty 2008). Through knockout of *Osx* in mice Nakashima *et al.* (2002) demonstrated that Osx functions downstream of RUNX2. There was no bone formation in Osx null mice, yet *Runx2* was still expressed. In addition, osteoblasts of these mice expressed chondrocyte markers such as *Sox9* and *Col2a1*. This therefore demonstrated that osterix is a negative regulator of chondrocyte cell specification and promotes osteogenesis.

1.1.3 Articular cartilage

Articulation of the vertebrate skeleton hinges on the function of synovial joints which form junctions between bones. The composition of synovial joints such as the knee allow efficient loading of weight, with the distribution of force across the articular

surface and joint cavity (Decker *et al.* 2014). Synovial joints are formed by several extracellular matrix rich tissues and cell types (Decker *et al.* 2015), as demonstrated in Figure 1.2. Briefly, the structure is enclosed by a joint capsule which is composed of fibrous tissue and forms a continuum of the periosteum. The cavity of the joint is lined with a synovial membrane and filled with synovial fluid (Rhee *et al.* 2005). Synovial fluid contains an array of molecules, including, growth factors, and lubricating factors such as hyaluronan and lubricin (Hui *et al.* 2012). This fluid sustains the avascular cartilage and tissues of the joint, provides lubrication and aids in the mechanical function of the joint. Within the knee, ligaments stabilise the structure and fibrocartilage forms the menisci which aid in the dispersal of mechanical loading across the joint (Lories and Luyten 2011; Salva and Merrill 2017). The bones that meet to form the joint are covered by hyaline articular cartilage, which is of critical importance in ensuring joint physiology (Goldring 2012).

1.1.3.1 Articular cartilage structure

Articular cartilage is split into three distinct zones in accordance with distance from the articular surface and properties of the chondrocytes that reside within; superficial, intermediate and deep (Heinegård and Saxne 2011). Superficial zone chondrocytes have flattened morphology and secrete matricellular proteins such as lubricin, that aid in the mechanics of the joint through reducing friction at the articular surface (Goldring 2012; Rhee *et al.* 2005). The intermediate and deep zone chondrocytes have rounder morphology and produce typical cartilage extracellular matrix components such as collagen type II (Decker *et al.* 2015). The extracellular matrix of articular cartilage is also highly organised, with composition dependent on localisation with respect to the chondrocyte cell, as demonstrated in Figure 1.2. The cell surface membrane of a chondrocyte contains many types of receptor that facilitate communication between the cell and ECM. Receptors for factors such as TGF-β and BMPs, as previously mentioned, allow transduction of signalling pathways that are key in chondrocyte behaviour (Kobayashi *et al.* 2005; van der Kraan and van den Berg 2012). The pericellular matrix is in the immediate vicinity to chondrocytes. This matrix contains proteins such as CCN2, perlecan and collagen type VI; but not collagen type II, and mediates cartilage biochemical and biomechanical structure in addition to signalling between the extracellular matrix and chondrocyte (Tang *et al.* 2018; Wilusz *et al.* 2014). Furthermore, the composition of this portion of the ECM prevents erroneous activation of

chondrocyte cell surface receptors. For example, the interaction between chondrocyte surface receptor discoidin domain receptor 2 (DDR2) and its ligand collagen type II is blocked by the pericellular matrix (Goldring 2012). Activation of DDR2 leads to MMP13 activity and therefore degradation of the cartilage matrix (Mackie *et al.* 2008). The territorial matrix surrounds the pericellular matrix and contains structural elements such as collagen type VI and aggrecan. The latter protein is also found in the inter-territorial matrix which is located most distally from the chondrocyte and forms a network with structural proteins such as collagen type II, IX and XI (Heinegård and Saxne 2011). This region is also marked by abundance of COMP (Zaucke *et al.* 2001). The organisation of the chondrocyte cell surfaces, extracellular matrix; and articular cartilage as a whole, enables efficient compression upon weight bearing and gliding movement with articulation.

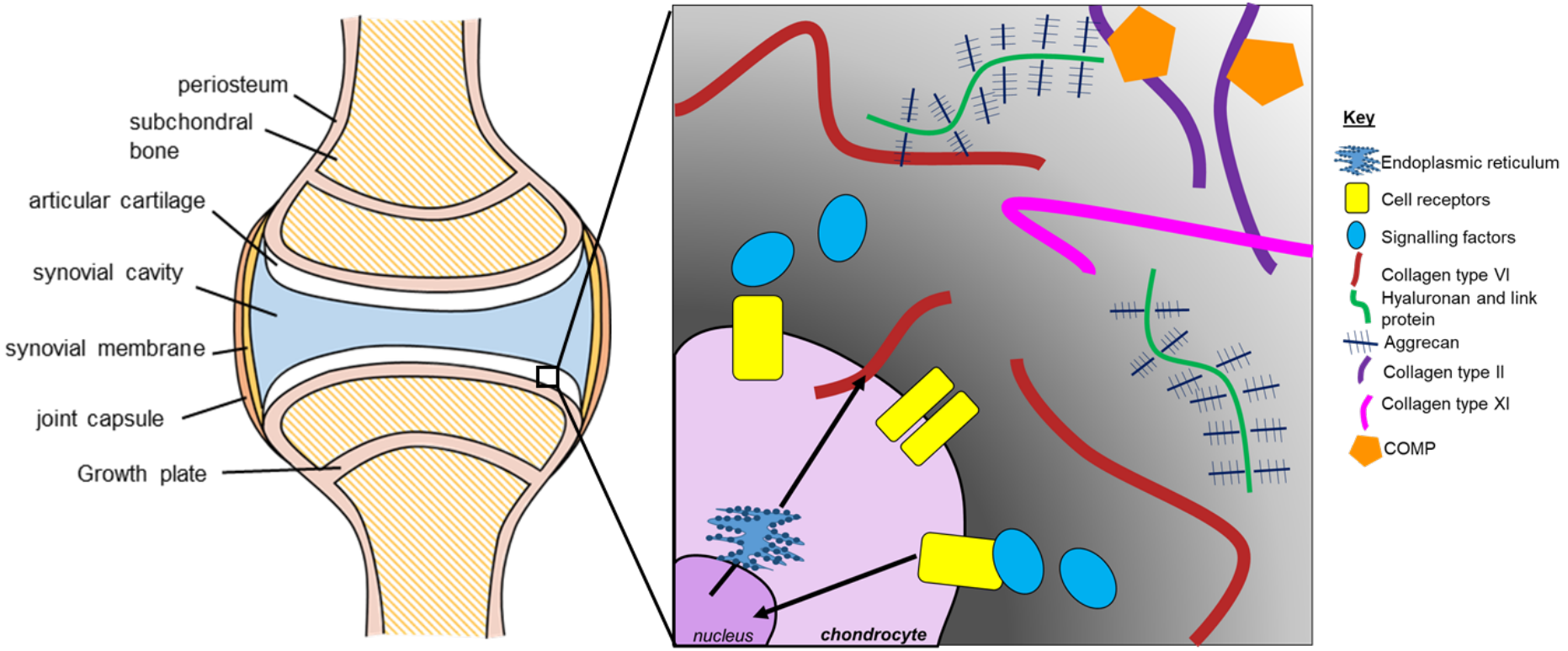


Figure 1.2: Schematic representation of a synovial joint and ECM secreted by articular chondrocytes. The epiphysis of two bones meet to form the joint which is enclosed by a fibrous joint capsule. The synovial joint cavity prevents direct contact between the two bones. The cavity is lined by a synovial membrane and filled with synovial fluid which lubricates the joint. Synovial fluid also prevents friction between the articular cartilage surfaces which line the bone epiphysis. This structure allows mechanical force to be distributed effectively across the joint. The metaphyseal growth plate is found below the subchondral bone and allows interstitial growth of the bone in the postnatal developmental period. The box shows organisation of cartilage extracellular matrix in greater detail. The pericellular matrix is enriched for cellular signalling molecules allowing efficient interaction with receptors on the surface of chondrocyte cells. The territorial matrix contains collagen type VI and aggrecan. This region is surrounded by the interterritorial matrix which is enriched for structural proteins such as collagen type II.

1.1.3.2 Articular cartilage development

Articular joints develop during embryogenesis from a mesenchymal interzone that forms between cartilaginous anlage at the site of the future joint (Decker *et al.* 2014; Koyama *et al.* 2008). Thus, chondrocytes within the interzone lose expression of chondrocyte markers including *Col2a1*, and instead express genes such as growth and differentiation factor 5 (*GDF5*), *Wnt9a* and *Noggin* (Koyama *et al.* 2008; Shwartz *et al.* 2016). In addition, this process is perpetuated through loss of *Sox9* expression in the interzone which is caused by Wnt signalling through β-catenin that promotes joint formation (Guo *et al.* 2004). *GDF5* expression marks synovial joint progenitor cells from E11.5 (Storm and Kingsley 1996). Segmentation of cartilage and formation of the synovial cavity is achieved through cells in the centre of the interzone undergoing apoptosis (Guo *et al.* 2004; Rhee *et al.* 2005). Cellular differentiation processes allow the assembly of joint tissues such as the synovial membrane and other structural features such as ligaments, and secretion of synovial fluid, therefore allowing joint function (Decker *et al.* 2014). Interzone cells have typically been viewed as progenitors for the synovial joint. For example, Koyama *et al.* (2008) used *GDF5* expression linked reporter gene based assay to demonstrate that interzone cells remained in the vicinity of the developing joint with localisation that was concomitant with that of articular chondrocytes upon joint formation. Therefore, articular chondrocytes were derived from interzone progenitors. (Koyama *et al.* 2008) The homogeneity of interzone cells in subsequent articular cartilage has, however recently been questioned. Shwartz *et al*. (2016) used a GDF5 lineage-tracing model, observing that articular joints surfaces were not exclusively populated with GDF5 positive cells. Therefore, not all of the joint-forming cells were derived from the interzone; indicating recruitment of cells from outside of the interzone. Moreover, this study also demonstrated that cells are continually recruited in order to form the joint. Recent work from Roelofs *et al*. (2017) has demonstrated that GDF5 expressing cells originating from the interzone are found in adult mouse synovium. Each of these studies highlight the complexity in the processes that determine cellular differentiation and stratification in tissue type during articular joint development.

1.1.3.3 Osteoarthritis

Postnatal regulation of articular chondrocyte behaviour is of integral importance in maintaining synovial joint physiology. Osteoarthritis (OA) is a profound and

currently incurable pathology caused by loss of synovial joint homeostasis (Chen *et al.* 2017). Changes in articular cartilage underpins aetiology, however the entire joint is affected in the disease (Goldring 2012). This progressive disease self-perpetuates with unpreventable and gradual destruction of the joint, which can cause severe pain as a result of friction within the joint. Several factors have been implicated in triggering OA, such as ageing, genetic susceptibility, injury and obesity (Mobasheri *et al.* 2017). Articular cartilage, menisci and ligaments degenerate and aberrant bone formation occurs with subchondral bone sclerosis and the appearance of osteophytes (Heinegård and Saxne 2011; Liu *et al.* 2018b). Chondrocytes within normal cartilage are quiescent, and the ECM that they secrete undergoes little turnover (Goldring 2012), thereby preserving cartilage homeostasis in addition to allowing efficient response to mechanical and molecular signalling events. This is exemplified by the fact that many of the ECM proteins within cartilage have extremely long half-lives in the order of decades (Mobasheri *et al.* 2017). This contrasts OA chondrocytes which exhibit increased proliferation with a greater rate of ECM component production (Lories and Luyten 2011). This triggers an increase in the expression and secretion of catabolic factors such as MMPs and a disintegrin and metalloproteinase with a thrombospondin motif (ADAMTS) which degrade the structural proteins of the ECM (Chen *et al.* 2017). For example, MMP13 degrades collagen type II; which is the most abundant form of collagen within articular cartilage (Goldring 2012). In a murine model with *MMP13* expression knockout, mice are protected from cartilage degradation after surgical induction of OA (Little *et al.* 2009). In addition, Echtermeyer *et al*. (2009) demonstrated that the syndecan 4 membrane receptor on the surface of chondrocytes is involved in osteoarthritis through enhanced ADAMTS-5 activity. Ultimately, loss of homeostasis within the joint favours catabolic activity. Inflammatory factors are expressed by several cell types within an OA joint including chondrocytes and synoviocytes (Goldring 2012). This therefore perpetuates the expression of catabolic enzymes and further cartilage degradation. Further change in articular chondrocytes is signified by switch to hypertrophic phenotype and expression of terminal differentiation markers including *Runx2* , in addition to ECM calcification (Lories and Luyten 2011). Changes in TGF-β signalling also contribute to OA, with switch from signalling through Smad2/3 which is chondroprotective, to Smad1/5/8 which promotes chondrocyte hypertrophy (van den Bosch *et al.* 2014).

1.2 CCN2 is an imperative protein within the matricellular environment

1.2.1 CCN2 is a member of the CCN family of proteins

The cellular communication network factor ((CCN) (Cyr61 (cysteine-rich protein 61), CTGF (connective tissue growth factor), NOV (nephroblastoma overexpressed gene)) family of proteins consists of six members (CCN1-6) (Perbal *et al.* 2018). These proteins function in extracellular signalling through interaction with matricellular signalling molecules and cell membrane receptors (Leask and Abraham 2006). CCN family members function in many biological processes such as cellular differentiation, migration, adhesion and proliferation within many tissues (Perbal 2004). The effects of the CCN family members may complement one another, or be antagonistic (Yeger and Perbal 2007). For example, CCN1 and CCN2 are pro-angiogenic, whereas CCN6 is anti-angiogenic (Perbal 2004). The family are cysteine rich and share protein structure which consists of four domains (modules I-IV) (Leask and Abraham 2006), as illustrated in Figure 1.3.

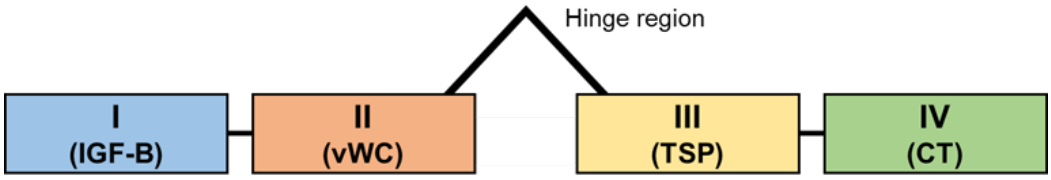


Figure 1.3: Modular structure of CCN family proteins underpins matricellular function Each CCN family member is composed of a modular structure containing domains (I-IV) that are capable of interacting with many signalling factors. The hinge region between modules II and III is sensitive to proteolysis (Krupska *et al.* 2015).

Each domain confers interaction with other proteins localised at cellular membranes and within the extracellular matrix (Leask and Abraham 2006). Module I consists of an insulin-like growth factor protein binding domain (IGFB), module II is a von Willebrand factor type C (vWC) binding domain which is linked to module III; a thrombospondin (TSP) domain, by a hinge region, and module IV is a C-terminal domain that contains a cysteine knot motif (Bork 1993). The peptide sequence of this region is variable between CCN family members (Holbourn *et al.* 2008). CCN5 does not contain the CT domain module IV (Perbal 2004). Many signalling factors interact with each domain, giving rise to the broad variety of processes in which CCN family members are involved. Moreover, CCN function may reflect simultaneous interactions of several signalling molecules across the domains of the CCN protein (Holbourn *et al.* 2008). This therefore allows the

integration of complex signalling pathways in a highly coordinated and specific manner (Jun and Lau 2011).

CCN2, systematically named as Cellular Communication Network Factor 2; also known as CTGF, hypertrophic chondrocyte-specific gene product-24 (HCS-24), Insulin-Like Growth Factor-Binding Protein 8 (IGFBP-8), FISP-12 or ecogenin is a fundamental mediator of matricellular signalling (Eguchi *et al.* 2008). The CCN2 protein is encoded by *Ccn2* (archaically annotated as *Ctgf*). Within the murine genome, *Ccn2* is located within the q arm of chromosome 10, within cytogenetic band A4 (10qA4). The gene is composed of five exons with four intervening introns; exon one codes for the N-terminal signal of the protein and the subsequent exons each encode the modules (I-IV) of the protein. First described by Bradham *et al*. (1991), within human umbilical vein endothelial cells, the 38kDa protein functions in many tissue types during development and beyond, including cartilage, skeletal, vascular, endocrine and neuronal tissues (Hall-Glenn and Lyons 2011). The modular structure of CCN2 mediates its activity. This was demonstrated by Grotendorst and Duncan (2005) who found that the C-terminal of the protein functions in the differentiation of myofibroblasts, whereas the N-terminal of the protein promotes fibroblast proliferation. Further studies have also shown that individual modules of the CCN2 protein mediate interaction with specific signalling factors and cellular proteins. Module IV of the CCN2 protein binds to heparin in adhesion of fibroblasts, myofibroblasts, endothelial and epithelial cells (Ball *et al.* 2003). CCN2 can directly bind to BMP4 and TGF-β through module II; inhibiting and enhancing the effect of these signalling pathways respectively (Abreu *et al.* 2002).

Friedrichsen *et al*. (2003) used *in situ* hybridisation methodology in order to examine the localisation of the mRNA product of *Ccn2* during embryonic development. *Ccn2* was strongly expressed within the vascular endothelium, perichondrium and neural tissue. This procedure was repeated by the same group in order to assess *Ccn2* transcription in adulthood (Friedrichsen *et al.* 2005). *Ccn2* was found to be expressed within mesenchymal tissue across several organs including the heart and kidney. However, aside from skeletal muscle, there is no mention of musculoskeletal system based *Ccn2* expression in adulthood, which is a major oversight given the expression of *Ccn2* in primitive forms of these tissues during embryonic development (Friedrichsen *et al.* 2003; Ivkovic *et al.* 2003). Given the extracellular nature of CCN2 protein function it is unsurprising that there may

be discrepancies between the localisation of its mRNA and protein product. For example, Nishida *et al*. (2003) found *Ccn2* mRNA product was localised within hypertrophic chondrocytes, yet CCN2 protein was found from the proliferative zone into the hypertrophic zone.

Aberrant expression of *CCN2* and *Ccn2* genes has been associated with several pathologies including fibrosis, cancer and osteoarthritis (Hall-Glenn and Lyons 2011). CCN2 has been well established as a mediator of fibrosis in many organs (Leask *et al.* 2009). Transgenic mice with fibroblast based overexpression of *Ccn2* display fibrosis in the skin, lung and kidney with enhanced proliferation of fibroblasts, differentiation into myofibroblasts and increased secretion of ECM factors such as fibronectin and tissue inhibitor of metalloproteinases 1 and 3 (TIMP-1 and 3) (Sonnylal *et al.* 2010). In a murine model of skin fibrosis, Mori *et al*. (1999) found that exogenous application of TGF-β and CCN2 potentiated fibrosis compared to treatment with only one of the factors. This synergistic interaction has subsequently been well documented in the perpetuation of fibrosis (Jun and Lau 2011). Ectopic expression of CCN2 has also been found in several cancers. Rachfal and Luquette (2004) found that CCN2 is expressed by several cell types within desmoplastic small round cell tumours, leading to the postulation that autocrine and paracrine CCN2 function further perpetuates tumorigenesis through enhanced angiogenesis and matrix deposition. The severity of breast cancer has been linked to the extent of CCN2 production within tumours (Xie *et al.* 2001). Furthermore, in an *in vivo* model of breast cancer metastasis, the capacity for metastatic progression was hampered through neutralisation of CCN2 function with an antibody (Shimo *et al.* 2006). This antibody against CCN2 was also used by Aikawa *et al*. (2006) in the attenuation of pancreatic cancer with decreased tumour growth, angiogenesis and cancer cell proliferation. Each of these studies therefore highlight the importance of stringent regulation of CCN2 expression and function.

1.2.2 CCN2 is fundamental in cartilage physiology

Each CCN family member is involved in skeletal development (Holbourn *et al.* 2008). CCN2 has been well documented as a major regulator of chondrocyte and cartilage physiology. The pivotal work of Ivkovic *et al*. (2003) revealed that global knockout of *CCN2* profoundly affects cartilaginous tissue. In this study, *Ccn2*$^{-/-}$ mice exhibited gross chondrodysplasia with dysmorphism in the chondrocranium, sternum and vertebrae, with kinked appearance of the ribs and long bones.

Aberration of rib development ultimately caused respiratory failure and perinatal mortality. Further scrutiny revealed that defects in stratification of the growth plate and resultant endochondral ossification were the cause of this severe phenotype. Firstly, expression of *Ccn2* in WT embryos was detected in the perichondrium at E12.5, with expression in hypertrophic chondrocytes increasing with the onset and progression of endochondral ossification. The role of CCN2 in the transition into hypertrophy was highlighted in the knockout mice as this is the zone where the long bones were misshapen, which suggests abnormality in the deposition of cells during the shift in differentiation state (Ivkovic *et al.* 2003). This in turn was shown to be due to reduced chondrocyte proliferation, consolidating a previous study by Nakanishi *et al*. (2000) who found that *in vitro*, *CCN2* overexpression in chondrosarcoma cells led to increased cellular proliferation. These findings are reiterated by those of Lambi *et al*. (2012) who described the proliferative zone being smaller, with enlarged hypertrophic zone in *Ccn2* null mice compared to WT control mice. Disruption of cartilage caused by the removal of *Ccn2* expression is further perpetuated by a loss of vital ECM components such as aggrecan (Ivkovic *et al.* 2003). CCN2 may also promote hypertrophy, with Maeda *et al*. (2009) describing interaction between BMP2 and CCN2 as leading to increased chondrocyte hypertrophy, reduction in cellular proliferation and increase in the expression of mature differentiation markers.

Hall-Glenn *et al*. (2013) also utilised *Ccn2*$^{-/-}$ mice in order to explore stress response in chondrocytes. Chondrocyte cell death was observed within the growth plate. In addition, cellular morphology, and more specifically the endoplasmic reticulum and nucleus, were altered compared to WT control. This was attributed to lack of interaction between the C-terminal of CCN2 and integrin α5β1 which endogenously aids in resistance to stress. *CCN2* null mice were also used by Maeda-Uematsu *et al*. (2014) who described the role of CCN2 in chondrocyte metabolism. Chondrocytes cultured *ex vivo* from knockout mice had decreased cellular adenosine triphosphate (ATP) level compared to wild-type control. The addition of recombinant CCN2 (rCCN2) to *Ccn2*$^{-/-}$ chondrocytes partially restored ATP concentration to that of the control. Furthermore, the expression of genes associated with energy metabolism and ATP production were reduced in knockout cells compared with control. Murase *et al*. (2016) built on this with the finding that amino acid and protein metabolism in *Ccn2*$^{-/-}$ mice was perturbed with a deficiency in amino acids. This led to the postulation that CCN2 interacts with and regulates free amino acids in cartilage ECM. Whilst chondrocytes are the primary cell type

affected in *Ccn2*$^{-/-}$ mice, Ivkovic *et al*. (2003) found that angiogenesis within the growth plate is also affected by loss of *Ccn2* expression. This aberration in angiogenesis reinforced abnormality in endochondral ossification, because vascular and osteoblast invasion, and therefore replacement of cartilage with bone was curtailed. However, on a global scale the vasculature was not perturbed by knockout of *Ccn2* (Ivkovic *et al.* 2003).

The overexpression of *Ccn2* also has ramifications in skeletal development. This has been explored in several mouse models harnessing conditional knock-in of *CCN2* expression in the context of other cartilage markers. Firstly, Nakanishi *et al*. (2001) generated transgenic mice overexpressing *Ccn2* in the context of collagen type XI, which as aforementioned, is a constituent of cartilage (Lefebvre and Smits 2005). In these transgenic mice, embryonic development of the skeleton occurred as in the wild-type control. This was surprising given that CCN2 is seen as promoting endochondral ossification, it would be expected that the transition to hypertrophy would occur prematurely, or chondrocytes would proliferate more rapidly prior to transition into hypertrophy, affecting interstitial growth embryonic limb length. However, in postnatal development mutant mice did display dwarfism and reduced mineralisation of bone compared to control. This was rationalised with the notion that premature angiogenesis and subsequent ossification caused by overexpression of *Ccn2* prevented growth in postnatal development rather than during embryonic development (Nakanishi *et al.* 2001).

The importance of considering the context of transgenic overexpression is highlighted by Tomita *et al*. (2013), who also used an approach of *Ccn2* overexpression within a cartilaginous context. This group used *Col2a1* linked overexpression, which led to transgenic mice with longer limbs at birth compared to wild-type. The effect of *Ccn2* overexpression within this context was therefore mediated during embryonic endochondral ossification, albeit no gross differences in skeletal structure were observed at E15.5. Nevertheless, later in embryonic development proliferation of chondrocytes was observed across the growth plate of transgenic embryos with increased bone length; whereas in wild-type controls proliferation was confined to the proliferative zone. By eight weeks of age, the femurs of transgenic mice were longer with greater mineralisation and cortical bone thickness than wild-type controls (Tomita *et al.* 2013). Mirroring the aforementioned work of Ivkovic *et al*. (2003), this study also found enhanced production of cartilage ECM components such as aggrecan and collagen type II in *Ccn2* overexpressing

mice. Moreover, chondrocyte proliferation was increased in these mice. Ultimately, the phenotype observed in these transgenic mice was attributed to increased rate of chondrogenesis and endochondral ossification (Tomita *et al.* 2013).

This transgenic mouse line with *Col2a1* based *Ccn2* overexpression was also used by Itoh *et al*. (2013) who examined the role of CCN2 in cartilage degeneration during adulthood. Overexpression of *Ccn2* in cartilage had a chondroprotective effect, limiting cartilage degradation associated with increasing age through a reduction in OA-like changes within the knee joints of transgenic mice compared with wild-type controls. *Ccn2* overexpression led to increased chondrocyte proliferation and cartilage ECM production, with repression of catabolic factors such as MMP13. Whilst this study suggests that CCN2 is important in maintaining cartilage physiology, the role of CCN2 in osteoarthritis is convoluted and unclear. Omoto *et al*. (2004) examined the expression of CCN2 in cartilage from OA patients, finding CCN2 was more abundant within OA cartilage and osteophyte tissue compared with unaffected control. This was postulated to be a contributory factor in the acquisition of a fibrotic phenotype in chondrocytes which perturbs cartilage homeostasis and leads to the reactionary instigation of catabolic factors that cause degradation and perpetuation of osteoarthritis. Nishida *et al*. (2004) induced OA in a rat model before addition of rCCN2 within a hydrogel matrix into the affected joint cavity. *Ccn2* expression was greater in cartilage in which OA had been induced. This therefore supported the consensus that CCN2 expression is induced in order to limit cartilage degradation. Recombinant CCN2 was observed to ameliorate cartilage defects triggered by OA, with rescue of normal chondrocytes residing in cartilage ECM; CCN2 therefore had a chondroprotective role (Nishida *et al.* 2004).

Abd El Kader *et al*. (2014) treated surgically induced cartilage injury with the module III TSP domain of CCN2 alone, observing that induction of cartilage repair was similar; if not better than the application of the complete CCN2 protein. This supports the principle that CCN2 is a positive influence in cartilage integrity and maintenance of physiology. Moreover, this study demonstrates how the modular structure of CCN2 mediates its function in cartilage. Nishida *et al*. (2003) applied rCCN2 protein to immortalised human chondrosarcoma (HCS) -2/8 cells, finding that the proteoglycan perlecan was co-localised with rCCN2. Furthermore, this interaction increased cellular proliferation. Aoyama *et al*. (2009) found that CCN2 is capable of direct interaction with aggrecan. More specifically, this was mediated

by modules I and II of CCN2. Module IV of CCN2 binds to fibronectin via integrin α5β1, promoting chondrocyte adhesion (Hoshijima *et al.* 2006). Interaction between CCN2 and fibronectin has also been implicated in mesenchymal condensation during chondrogenesis (Song *et al.* 2007). Khattab *et al*. (2015) found that rCCN2 interacts with other cartilaginous ECM molecules including GDF5, VEGF and TGF-β, but not PTHrP. However, this study did not determine the points within the CCN2 protein that facilitate these interactions. Further work is required in order to understand the intricacies of these interactions and their effect in cartilage.

The relationship between CCN2 and TGF-β has been well-documented, and this interaction also occurs in chondrogenesis and cartilaginous tissue. For example, Song *et al*. (2007) used an *in vitro* approach to demonstrate that CCN2 functions during the condensation of mesenchymal cells, and therefore in the early stages of chondrogenesis and endochondral ossification. This group utilised a well-established model of preliminary chondrogenesis consisting of application of TGF-β to immortalised C3H10T1/2 mesenchymal cells. In this study CCN2 expression and function was induced by TGF-β and contributed to the process of chondrogenesis (Song *et al.* 2007). The importance of this is unclear as Ivkovic *et al*. (2003) did not find aberration in initial chondrogenesis in the *CCN2*$^{-/-}$ mice. A recent study by Tang *et al*. (2018) found that knockout of *Ccn2* in cartilage in a post-natal manner led to reduced susceptibility to OA, with thicker cartilage; contradicting the consensus of CCN2 as being chondroprotective. This study also suggested that CCN2 controls the function of TGF-β in cartilage through direct interaction between the proteins that control the release of TGF-β. This is a controversial result as it suggests that CCN2 functions in the latent complex of proteins that sequesters TGF-β and regulates function (Tang *et al.* 2018). Furthermore, in this study low-density lipoprotein receptor-related protein 1 (LRP1) did not interact with CCN2 in regulating function; contradicting previous work that has characterised interaction between the two proteins in chondrocytes (Segarini *et al.* 2001). LRP1 is bound by many factors; both at the extracellular surface and within the intracellular space, thereby mediating cellular signalling, lipid metabolism and endocytosis (Herz and Strickland 2001). Kawata *et al*. (2006) described LRP1 as a receptor for CCN2 in chondrocytes, with LRP1 aiding the endocytosis of CCN2. The same group examined this interaction further, finding that LRP1 facilitates transcytosis of CCN2 in chondrocytes. Moreover, transcytosis has been posited as a method of distributing CCN2 protein across zones of the growth plate

from cells expressing *Ccn2* such as pre-hypertrophic chondrocytes (Kawata *et al.* 2006).

1.2.3 CCN2 function within osseous tissue

Whilst the function of CCN2 in chondrocytes and cartilage has been widely explored, CCN2 also has key functions in bone. Safadi *et al*. (2003) demonstrated that *CCN2* mRNA and protein product are expressed in rat bone at two weeks of age, with *Ccn2* mRNA predominantly found in the osteoblasts lining the trabeculae. In this study, and others CCN2 protein has been found in active osteoblasts and in some osteocytes (Kawaki *et al.* 2011; Safadi *et al.* 2003). Further results from the work of Safadi *et al*. (2003) suggest that CCN2 functions in osteoblast proliferation and differentiation, in addition to matrix production and mineralisation. CCN2 has also been shown to function in osteoclastogenesis, potentiating the differentiation of osteoclasts from haemopoietic precursors (Nishida *et al.* 2011).

Several studies have gone on to further characterise the role of CCN2 within bone using transgenic murine models. In *Ccn2*$^{-/-}$ mice, mineralisation of the femora was decreased, with reduced trabecular bone formation that was attributed to defective angiogenesis rather than osteoblastic cell function (Ivkovic *et al.* 2003). Kawaki *et al*. (2008) investigated intramembranous ossification in *Ccn2*$^{-/-}$ mice finding that osteoblast proliferation was reduced in null mice, with rescue of cellular proliferation in a dose dependent manner with rCCN2 treatment in osteoblast cultures. Expression of *Col1a1* and Bone gamma-carboxyglutamate protein (*BGLAP*) were reduced in *Ccn2* null calvarial osteoblasts and bone mineralisation was also reduced in knockout mice compared to wild-type control (Kawaki *et al.* 2008). Lambi *et al*. (2012) further characterised the skeletal impact of the removal of *Ccn2* expression and demonstrated that bone formation and the expression of *Runx2*, alkaline phosphatase (*ALP*), and *BGLAP* were increased in the kinked diaphysis of *Ccn2* null mice. However, this trend was reversed in the parietal bone. In addition, the trabecular bone of the femoral metaphysis of knockout mice was reduced (Lambi *et al.* 2012). This discrepancy in bone formation could be accounted for by the different regulatory signalling interactions that control endochondral and intramembranous ossification, and their effect on *Ccn2* expression. Yamaai *et al*. (2005) found that knockout of *Runx2* led to decrease in CCN2 abundance within skeletal elements during embryonic development and at postnatal day 0 (P0). Overexpression of *Ccn2* within an osseous context also causes defective

skeletogenesis. Smerdel-Ramoya *et al*. (2008) generated transgenic mice over-expressing *Ccn2* within the context of *BGLAP* expression. This study found that overexpression of *Ccn2* led to decreased bone mineral density and apposition, with decreased rate of bone formation and trabecular bone volume compared to WT control (Smerdel-Ramoya *et al.* 2008).

These studies therefore highlight the importance of tight regulation of *Ccn2* within osseous tissue in order to ensure physiological bone volume and density. Moreover, these studies suggest that CCN2 may function in feedback loops within osseous tissue in a spatially specific manner. Luo *et al*. (2004) used an *in vitro* model of osteogenesis to demonstrate that CCN2 acts downstream of Wnt3A and BMP9 during the early differentiation period. This study also showed that the effect of Wnt3A was dependent on canonical Wnt signalling, and that the ultimate effect of BMP9 signalling was dependent on CCN2 expression, yet overexpression of CCN2 repressed BMP9 and Wnt3A function. Treatment of osteoblasts with BMP2 has also been found to increase the expression of *Ccn2* (Parisi *et al.* 2006). In addition, Arnott *et al*. (2007) stimulated osteoblasts with TGF-β and found an increase in the production of CCN2 mRNA and protein. Canonical notch signalling also regulates *CCN2* expression in osteoblasts (Canalis *et al.* 2014). CCN2 function within bone is therefore governed by complex cellular signalling mechanisms and regulatory interactions; much of which remains unknown.

1.3 Enhancers mediate temporospatial specificity in gene expression

1.3.1 Gene transcription underpins every biological process

The central dogma of molecular biology dictates that deoxyribonucleic acid (DNA) sequence is firstly read and transcribed into a ribonucleic acid (RNA) intermediate before translation into amino acid sequence and assembly into functional protein (Crick 1970). Whilst a vast proportion of the genome is transcribed, protein coding genes only reflect a small fraction of transcriptional template (Djebali *et al.* 2012; The ENCODE Project Consortium 2012). The production of each transcript must be stringently regulated in order to prevent aberrant expression that would ultimately lead to pathology. The holoenzyme RNA polymerase II (Pol II) forms the basis of the protein machinery required for the interpretation of DNA sequence and catalysis of complementary RNA (Sayre *et al.* 1992). Transcription occurs in bursts with cycling of key stages; initiation, pausing, elongation and termination (Bartman

et al. 2016). The rate at which Pol II is recruited to the promoter and moves along the DNA template catalysing RNA production, in addition to the number of transcripts produced per burst are important factors in the efficiency of transcription (Fukaya *et al.* 2016). Transcription begins, and is primarily regulated at the non-coding regulatory DNA sequence upstream of a gene; its promoter region. The core promoter is the segment of this locus most proximal to the transcription start site (TSS) and typically spans -40 base pairs (bp) to +40bp of this site (Juven-Gershon and Kadonaga 2010). Core promoters are typically classed as being 'focused' or 'dispersed' based on the positioning of TSS and transcription factor binding sites; which results in corresponding transcription initiation patterns. Focused transcription is directed from specific TSS and frequently occurs for genes that are expressed in a highly cell-specific manner (Haberle and Stark 2018). This contrasts with dispersed core promoters from which transcription is initiated from several TSS, which is favoured in the expression of many housekeeping genes (Juven-Gershon and Kadonaga 2010). Regardless of the shape of a core promoter, it contains TFBS for the protein machinery and accessory proteins that are required for gene transcription (Levine *et al.* 2014). General transcription factors (GTF); transcription factors II A, B, D, E, F and H, along with TATA binding protein (TBP) interact with this region using motifs such as the TATA box, TFIIB recognition element (BRE), initiator (Inr) and downstream promoter element (DPE) (Juven-Gershon and Kadonaga 2010; Roy and Singer 2015). Interactions with these factors in addition to co-activator proteins and chromatin remodelling proteins, in conjunction with the recruitment of Pol II allows formation of the classical pre-initiation complex (PIC), although the proteins involved in this are highly variable (Levine *et al.* 2014) (illustrated in Figure 1.4).

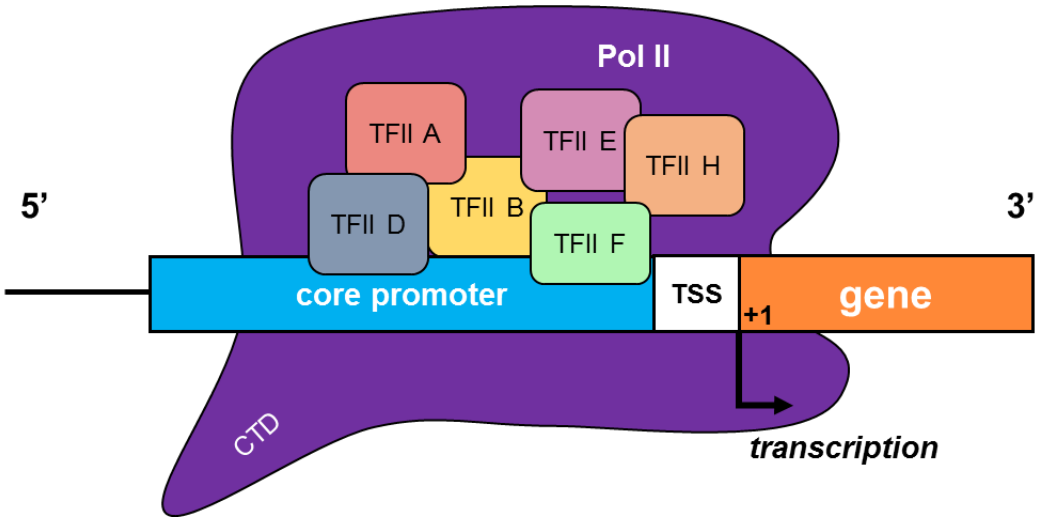


Figure 1.4: Assembly of the transcription pre-initiation complex at the core promoter. General transcription factors bind to consensus recognition binding motifs within the core promoter upstream of the gene (5' direction). RNA polymerase II is recruited to the core promoter and the pre-initiation complex is formed with general transcription factors II A, B, D, E, F and H. Phosphorylation of the C-terminal domain (CTD) of Pol II enables release from the promoter and movement along the coding sequence and therefore transcription in the 3' direction from the transcription start site (TSS).

The proximal promoter is the regulatory locus that resides further upstream of the TSS than the core promoter. This region contains binding sites for cell lineage-specific transcription factors that contribute to basal rates of gene expression (Harmston and Lenhard 2013).Transcription is initiated by phosphorylation of the serine residues within the C-terminal domain of Pol II, which allows release of the enzyme from the TSS. Pol II subsequently moves along the coding sequence catalysing the production of complementary mRNA. The progress of transcription is halted as Pol II stalls approximately 50bp downstream of the TSS (Core *et al.* 2012; Muse *et al.* 2007). This pausing confers an essential regulatory step in transcription (Gilchrist *et al.* 2010; Haberle and Stark 2018; Meng and Bartholomew 2018). Pausing is mediated by interaction between Pol II and negative elongation factor complex (NELF), with release of Pol II and the elongation phase of transcription triggered by interaction with positive elongation factor b (P-TEFb) (Core *et al.* 2014). Pol II accumulates at the promoter and is stabilised by the process of pausing, therefore allowing greater organisation and efficiency in the production of RNA transcript (Core *et al.* 2012; Henriques *et al.* 2018). Pausing has been posited as a facilitator of rapid transcription upon a regulatory cue that enables Pol II release, or through interaction between paused Pol II and other regulatory proteins that boost transcription rates (Lagha *et al.* 2012). Furthermore, promoters at which paused Pol II is observed ultimately exhibit greater transcriptional output (Core *et al.* 2008). The process of transcription is terminated

as RNA pol II reaches the end of the template sequence (terminator sequence) with change in protein affinity interaction with the template triggering detachment of the polymerase, in addition to the mRNA transcript detachment from Pol II (Proudfoot 2016).

Transcription typically occurs in a bidirectional manner (Meng and Bartholomew 2018). Core *et al*. (2008) found that approximately 80% of active genes in IMR90 human lung fibroblast cells demonstrated Pol II occupancy upstream of the TSS, resulting in RNA production upstream of the TSS. However, mRNA is only produced from downstream read-through of coding sequence. Scruggs *et al*. (2015) built on this model with the identification of multiple discrete TSS within a single promoter that facilitate bidirectional transcription, with the distance between these loci impacting on promoter structure leading to the postulation that this confers further regulation of gene transcription. Upstream antisense RNAs (uaRNAs) that result from divergent transcription are unstable and therefore rapidly degraded (Haberle and Stark 2018).

1.3.2 Gene transcription is dependent on chromatin state

The behaviour of any cell depends on the highly specific profile of cell lineage-specific genes that it expresses at any given time (De Laat and Duboule 2013). Regulation of gene transcription is therefore of paramount importance in cellular identity and tissue homeostasis. A pivotal aspect of this control is modulation of the interactions between DNA sequences and corresponding regulatory elements (Stampfel *et al.* 2015). Within the nucleus DNA is packaged as chromatin, of which there are two forms; euchromatin and heterochromatin (Gibcus and Dekker 2013). Euchromatin has an open structure that allows interaction between DNA sequence and regulatory elements and therefore gene transcription. This contrasts heterochromatin which has a highly compacted structure that inhibits gene expression. Chromatin is spatially isolated within the nucleus on the basis of these two states (Gibcus and Dekker 2013). Therefore, chromatin is highly dynamic and state is dependent on cellular context, which in turn reinforces cell-specific patterns of gene transcription (Ernst *et al.* 2011; Gibcus and Dekker 2013). Consequently gene expression can modify chromatin state and therefore 3D organisation of the genome (Rowley *et al.* 2017).

Chromatin is highly ordered with hierarchical structure. Gross scale organisation occurs with the formation of chromosomes, which occupy specific areas; chromosome territories, within the nucleus (Petit *et al.* 2017). Within chromosomes, chromatin is further organised on the basis of activity with spatial separation into A/B compartments (Lieberman-Aiden *et al.* 2009). A compartments have an open structure, containing many genes which are transcribed, whereas B compartments are transcriptionally silent with a compacted structure (Fortin and Hansen 2015; Gibcus and Dekker 2013). Chromosomal compartments are then divided into topologically associated domains (TADs) as illustrated in Figure 1.5 (Bonev and Cavalli 2016). These regions encompass several genes and cognate regulatory sequences and are largely conserved between cell types and organisms (Dixon *et al.* 2012; Gong *et al.* 2018b). Using Hi-C chromatin conformation capture based methodology, Dixon *et al*. (2012) initially described TADs as having a typical span of approximately 880 kilobases (kb), however subsequent refinement of this technique has led to the consensus that TAD size may vary greatly, yet on average is approximately 185kb (Bonev and Cavalli 2016; Rao *et al.* 2014). TAD function hinges on 3D chromatin organisation which facilitates interaction between DNA sequences. Furthermore, intra-TAD interaction is a major aspect of cell-specific gene transcription, with dynamism in these interactions mediating regulation of transcription (Fukaya *et al.* 2016; Plank and Dean 2014).

The boundaries of TADs are enriched for architectural protein motifs; namely CCCTC-binding factor (CTCF) and cohesin (Dixon *et al.* 2012). These proteins are frequently found in close proximity to one another (Tang *et al.* 2015). Whilst classically viewed as a repressor of transcription through insulator function, CTCF is increasingly being hailed as a facilitator of gene transcription (Ong and Corces 2014). CTCF has a major role in organising chromatin topology and function, with Guo *et al.* (2015) demonstrating that genome editing based disruption of CTCF sites perturbs chromatin topology. This is reinforced by the work of Nora *et al*. (2017) who showed that deletion of CTCF affects the looping within a TAD between the boundaries and that CTCF therefore mediates TAD insulation. Cohesin has classically been associated with chromosome structure, but also functions in organisation of chromatin topology. This protein is pre-requisite in TAD formation and stabilises CTCF interactions (Nuebler *et al.* 2018; Ong and Corces 2014). Busslinger *et al*. (2017) demonstrated that cohesin is unable to interact at TAD boundaries when CTCF is removed. Whilst the studies of Schwarzer *et al*. (2017) and Rao *et al*. (2017) showed that depletion of cohesin and its capacity to interact

with chromatin causes loss of TAD arrangement and chromatin looping. These findings highlight the importance of interaction between CTCF and cohesin in the efficient organisation of chromatin and subsequent regulation of gene expression. TAD boundaries prevent non-specific inter-TAD interactions between regulatory elements and therefore reinforce fidelity in interactions between regulatory elements and cognate genes, thereby repressing aberrant gene transcription (Figure 1.5B). For example, Lupiáñez *et al*. (2015) used clustered regularly interspersed short palindromic repeats- Caspase 9 (CRISPR-Cas9) genome editing to manipulate the boundaries of a genomic locus in which structural variants are associated with pathological limb phenotypes in humans. Disruption of the TAD resulted in loss of specificity in regulatory element to cognate gene interactions, leading to aberrant gene expression and recapitulation of disease phenotype in murine models. The strength of isolation provided by TAD boundaries is variable, with greater binding of CTCF denoting a stronger boundary (Gong *et al.* 2018b).

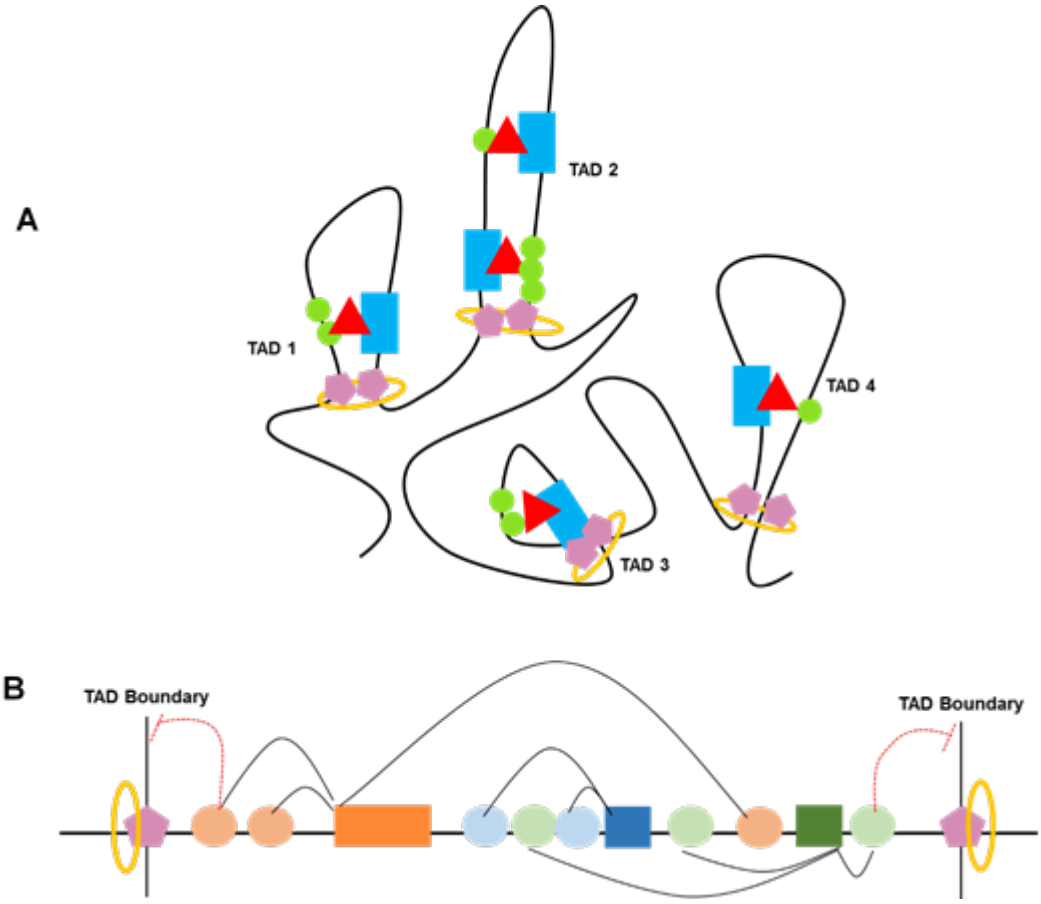


Figure 1.5: Chromatin is segregated into topologically associated domains (TADs). Chromosome compartments may contain several TADs (A). Chromatin is looped through the function of structural proteins CTCF (purple pentagon) and cohesin (yellow loop). This enables genes (blue rectangle, A) regulatory elements (green circles, A) and protein machinery required for transcription to be brought together. Within a TAD (B) there may be several genes (rectangles) and their cognate regulatory elements (circles). TAD boundaries are marked by cohesin and CTCF, enabling intra-TAD looping and preventing non-specific inter-TAD interaction.

The basic unit of chromatin is a nucleosome, which is composed of an octamer of histone proteins (two each of histone proteins H2A, H2B, H3 and H4), around which 147bp of DNA is wound (Richmond and Davey 2003). Nucleosome positioning and occupancy varies across the genome and is a factor in DNA accessibility (Struhl and Segal 2013). Nucleosome depleted regions (NDR) occur in specific loci across the genome, such as in regulatory elements (Scruggs *et al.* 2015; Struhl and Segal 2013). For example, nucleosomes are often absent at core promoters where transcription is occurring (Haberle and Stark 2018). Chromatin remodelling enzymes facilitate nucleosome sliding which causes shifts in nucleosome positioning and therefore DNA availability for interaction with regulatory proteins (Ho and Crabtree 2010). Post-translational modifications of histone proteins have a major role in chromatin state. A plethora of chromatin remodelling reader and writer enzymes are capable of recognising and catalysing variable modification of the histones such as methylation and acetylation, which impact on nucleosome compaction and DNA accessibility (Calo and Wysocka 2013; Heintzman *et al.* 2007). For example, trimethylation of lysine 27 residue of histone 3 (H3K27me3) is associated with repressive chromatin state. This modification is deposited by Polycomb repressive complex 2 (PRC2), leading to increased chromatin compaction and therefore silencing of gene expression (Margueron and Reinberg 2011). This contrasts acetylation modification of the same residue (H3K27ac) which characterises regulatory regions associated with active transcription (Kieffer-Kwon *et al.* 2013; Rada-Iglesias *et al.* 2011). Moreover, Stasevich *et al*. (2014) found that H3K27ac aids Pol II in the induction of the elongation step of transcription. Antagonism between H3K27ac and H3K27me3 is therefore important in regulation of chromatin state. In addition, this highlights the fact that regulatory element function within the genome can be identified on the basis of histone modification.

1.3.3 Enhancers are fundamental regulators of gene transcription

Whilst the promoter region for a gene orchestrates basal levels of gene transcription, further regulatory elements are required to refine the process of gene transcription (Zabidi and Stark 2016). Enhancers are an imperative class of *cis*-regulatory module (CRM) that mediate transcription of a target gene in a highly specific temporospatial manner. Recent advances in high-throughput sequencing methodology has galvanised the study of enhancers, for example The Encyclopaedia of DNA Elements (ENCODE) Project Consortium (2012) identified

399,124 loci within the human genome that had enhancer-like characteristics. Given the latest estimate that the human genome contains approximately 21,000 protein coding genes, the number of putative enhancers demonstrates the complicated regulatory mechanisms that allow the transcriptional regulation of each gene in every cell throughout life (Pertea *et al.* 2018). There may be several enhancers for a single gene, with each functioning in a specific cellular context (Kieffer-Kwon *et al.* 2013). Furthermore, many enhancers exhibit function that is specific to one cell type (Ernst *et al.* 2011). The importance of these elements is exemplified by the fact that the transcriptional output from a promoter alone is weak when compared with the combinatory action of a promoter and enhancer (Shlyueva *et al.* 2014).

1.3.3.1 Enhancer function is dependent on interaction with promoter regions

Interaction between an enhancer and the promoter region of its target gene underpins enhancer activity (Li *et al.* 2012). TAD organisation of chromatin is critical in facilitating this; highlighting a key characteristic of enhancers, the ability to interact with their target gene regardless of distance or orientation (Haberle and Stark 2018; Harmston and Lenhard 2013). There may be several genes and their cognate enhancers within a TAD, therefore interaction between an enhancer and target gene promoter must be highly specific in order to ensure efficient transcription (Zabidi and Stark 2016). TADs may be divided into 'sub-TAD' regions allowing the segregation of genes that are being actively transcribed and their cognate enhancers (Dowen *et al.* 2014). Enhancers are irregularly activated with expression of non-target genes upon disruption of CTCF and cohesin binding at sub-TAD boundaries (Hanssen *et al.* 2017). Dynamism in chromatin structure facilitates specificity in function, as interactions are dependent on cellular differentiation state (Rubin *et al.* 2017; Zhu *et al.* 2013a). For example, the work of Dowen *et al*. (2014) and later Hanssen *et al*. (2017) demonstrated that TAD and sub-TAD organisation enables regulatory elements to be activated or repressed through a common mechanism of insulation, in a cell-specific and tissue-specific manner. The extent of stringency in the interaction between an enhancer and corresponding promoter can vary with cell type, which correlates with gene function; for example there is less stringency in broad functioning housekeeping enhancer interactions compared with those of cell lineage-specific enhancers (Zabidi *et al.* 2015).

There are multiple models as to how the promoter and enhancer communicate, but the 'looping model' has strongest consensus as to how these two genomic regions are able to interact (Harmston and Lenhard 2013). This model dictates that the enhancer and promoter are brought together through a hierarchy of interactions that causes the intervening chromatin between the two regulatory loci to become looped (Harmston and Lenhard 2013; Meng and Bartholomew 2018). Enhancer-promoter communication is dependent on an array of DNA-protein and protein-protein interactions. Enhancers are formed of clusters of TFBS, which are utilised through interaction with cell lineage-specific combinations of transcription factors (Calo and Wysocka 2013; Reiter *et al.* 2017). The advent of Chromatin immunoprecipitation Sequencing (ChIP-Seq) methodology has enabled genome wide interrogation of TFBS; a useful tool in the identification of putative enhancer regions (Dogan *et al.* 2015; Visel *et al.* 2009a). The binding of 'pioneer' transcription factors to enhancer regions is a preliminary step in the commissioning of enhancer function (Spitz and Furlong 2012). Pioneer factors are transcription factors that are capable of interacting with inactive, closed chromatin (Iwafuchi-Doi and Zaret 2014). Cell lineage specificity in the function of these TF reinforces temporospatial specificity in enhancer function (Donaghey *et al.* 2018). Pioneer factor interaction facilitates the recruitment of further transcription factors and chromatin remodelling proteins (CR) (Reiter *et al.* 2017). More specifically, the binding of pioneer factors causes nucleosome displacement and reorganisation that results in open chromatin structure that is capable of interaction with further regulatory proteins; therefore mediating enhancer function (Iwafuchi-Doi and Zaret 2014; Shlyueva *et al.* 2014).

The spacing, orientation, order and number of TFBS within an enhancer dictate transcription factor interaction; a concept known as 'enhancer grammar' (Harmston and Lenhard 2013). There are several models of enhancer grammar and the mechanisms through which transcription factors assemble at enhancers. Firstly, the 'enhanceosome' model denotes enhancers that consist of strictly arranged TFBS, with enhancer output dependent on cooperative binding between highly ordered transcription factors positioned relative to one another (Spitz and Furlong 2012). In addition, the function of enhancers displaying this grammar is more susceptible to mutations which may cause the loss of TFBS and therefore transcription factor interaction, halting enhancer activity (Dickel *et al.* 2013). This contrasts the flexibility exhibited in the 'billboard' model. First proposed by Kulkarni and Arnosti (2003), this model is based on enhancers composed of clusters of

TFBS positioned in a non-specific order, with enhancer activity reflecting the input of multiple TFs that may have antagonistic impact on enhancer function. Transcription factor binding occurs on an independent basis and enhancer function is not dependent on cooperative and sequential interaction of these proteins (Harmston and Lenhard 2013). Furthermore, in this form of organisation the type of transcription factor bound is more important than where it positioned (Long *et al.* 2016). Given the stringency of the enhanceosome organisation, it is not surprising that is not observed as commonly as the more robust billboard model (Reiter *et al.* 2017). A further model, the TF-collective model has been used to describe transcription factor binding at enhancers where TF interact indirectly with enhancer motifs through protein-protein interactions (Long *et al.* 2016; Spitz and Furlong 2012). The complexity in transcription factor binding at enhancers is compounded by the fact that many enhancers exhibit characteristics of all models (Dickel *et al.* 2013; Long *et al.* 2016). For example, a recent study by Grossman *et al*. (2018) demonstrated that the motifs for specific groups of TF lie within particular regions of the enhancer. This further reinforces the importance of TF binding organisation across an enhancer, yet emphasises the fact that the extent and fidelity of transcription factor binding interaction in driving enhancer function is still very much unknown.

Further hierarchical interactions between transcription factors bound to the enhancer, structural proteins, transcriptional co-activators and chromatin remodelling factors such the histone acetyltransferases cAMP response element-binding protein (CREB) binding protein (CBP) and p300, culminates in the enhancer-promoter interaction that underlies enhancer activity (Grossman *et al.* 2018; Visel *et al.* 2009a). The Mediator complex is an important aspect of this and has classically been viewed as a 'bridge' between the two regulatory loci (Bonev and Cavalli 2016; Yin and Wang 2014). This complex is formed of approximately 30 protein subunits which function in the coordination of signals between enhancer and promoter; an individual Mediator complex interacts with both enhancer and promoter (Petrenko *et al.* 2016; Soutourina 2017). The seminal work of Kagey *et al*. (2010) demonstrated that Mediator interacts in a cell-specific manner alongside cohesin in order to facilitate association between an enhancer and promoter. Moreover, Mediator can contribute to the release of Pol II in transcription initiation (Petrenko *et al.* 2016). As aforementioned, CTCF is a critical regulator of chromatin looping and is therefore an important aspect of enhancer-promoter communication, and can be bound to both elements (Ong and Corces 2014; Weintraub *et al.* 2017).

Ren *et al*. (2017) demonstrated that the binding of CTCF to motifs in close proximity to enhancers facilitates the proximity between interacting enhancers and promoters, in addition to stabilising the interaction between these two regions. Yin Yang 1 (YY1) is a further protein that has been associated with enhancer-promoter interaction, through binding at both elements (Weintraub *et al.* 2017).

Enhancer-promoter interaction and therefore enhancer function positively influences transcriptional bursting, which results in increased gene transcription (Catarino and Stark 2018). For example, strong enhancers are associated with greater burst frequency (Fukaya *et al.* 2016). However, the extent of enhancer interaction throughout the entirety of a transcriptional burst is not yet known (Bartman *et al.* 2016). The strongest consensus on enhancer function is that enhancer-promoter interaction aids in the recruitment and stabilisation of the PIC at the promoter region and therefore boosts transcription initiation rate (Haberle and Stark 2018; Meng and Bartholomew 2018). A schematic representation of these interactions is outlined in Figure 1.6.

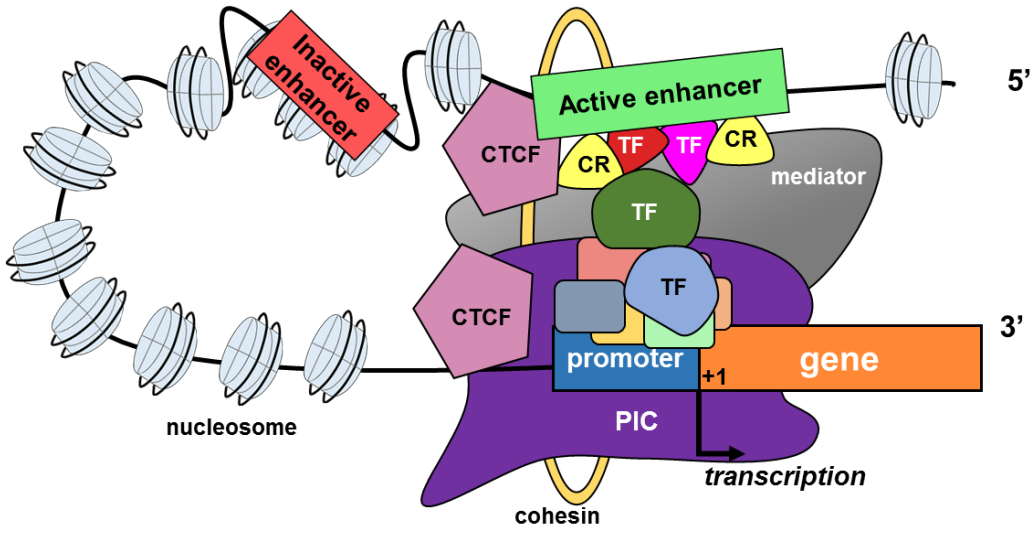


Figure 1.6: Enhancer-promoter interaction facilitates recruitment of the pre-initiation complex (PIC) to the promoter and the initiation of gene transcription. Cell lineage-specific pioneer factors and co-activator transcription factors (TF) interact with both the active enhancer and promoter regions. Subsequent interactions with chromatin remodelling proteins (CR), structural proteins including CTCF, cohesin and Mediator allow chromatin looping with the enhancer and promoter being brought into close proximity with one another. This culminates in the recruitment of the PIC to the gene promoter and the initiation of gene transcription. Inactive enhancers are spatially unable to interact with the target promoter through chromatin looping.

Enhancers are also capable of influencing transcription at the elongation stage. Interaction between enhancers and regulatory proteins can aid in the recruitment of p-TEFb complex and the release of Pol II from the transcriptional pausing stage

(Haberle and Stark 2018). Liu *et al*. (2013) described 'anti-pause enhancers' as enhancers that regulate the release of Pol II from pausing stage and trigger transcriptional elongation. An influential study by Sawado *et al*. (2003) demonstrated that deletion of the locus control region enhancer for the β-globin gene caused a decrease in pre-initiation complex assembly, in addition to reduced phosphorylation of Pol II and transition into elongation phase; this enhancer therefore functioned in multiple stages of gene transcription.

Whilst enhancers function as regulators, they must in turn be regulated in order to ensure stringent specificity in patterns of gene transcription (Visel *et al.* 2009b). Chromatin state is an integral factor in enhancer activity. As outlined previously, chromatin with closed and compacted structure is unable to interact with regulatory proteins, enhancers subject to this chromatin state are inactive (Heinz *et al.* 2015). As aforementioned, pioneer factors are able to overcome inactive chromatin state and activate enhancer activity, which is an integral stage in cellular reprogramming and differentiation (Iwafuchi-Doi and Zaret 2014). Enhancers often exhibit nucleosome depletion as a result of pioneer factor binding, with open chromatin state causing hypersensitivity to deoxyribonuclease I (DNase I), which is an attribute that is commonly used to predict enhancer regions (Heinz *et al.* 2015; Shlyueva *et al.* 2014; Thurman *et al.* 2012). Where nucleosomes are present in an enhancer, histones may be substituted with non-canonical histone variants such as H2A.Z and H3.3, the former of which is associated with strong enhancers (Calo and Wysocka 2013; Ernst *et al.* 2011; Meng and Bartholomew 2018). These histone variants cause greater instability in nucleosome structure, which allows them to be displaced from DNA more easily, enabling greater rates of TF-DNA interaction (Harmston and Lenhard 2013). DNA methylation state also influences the utilisation of enhancers; active enhancers are often hypomethylated (Kieffer-Kwon *et al.* 2013). Sheaffer *et al*. (2014) found that depletion of the expression of DNA methyltransferase 1 (DNMT1); which maintains DNA methylation, led to aberrant activation of gene transcription through pathological enhancer-TF interaction which was repressed by DNA methylation in wild-type samples. (Aran and Hellman 2013). King *et al*. (2016) described DNMT1 based DNA methylation in the regulation of histone methylation which therefore reinforces regulation of enhancer activity. In addition, abnormal methylation of enhancer DNA has been associated with cancer (Fleischer *et al.* 2017).

1.3.3.2 Posttranslational modification of histones affects enhancer activity

The post-translational modification of histones is an important factor in the identification of putative enhancer regions and can be used to predict the activity of an enhancer at a given time point (Calo and Wysocka 2013; Rada-Iglesias *et al.* 2011; Zhu *et al.* 2013b). The work of Heintzman *et al*. (2007) demonstrated that enhancers can be distinguished from promoters due to the extent of methylation of the fourth lysine of histone protein three; promoters are marked with trimethylation of this residue (H3K4me3), whereas enhancers are associated with monomethylation of the residue (H3K4me1). Moreover, H3K4me1 is intrinsic at enhancer regions (Bonn *et al.* 2012). As previously mentioned, H3K27ac is another histone modification associated with regulatory function which can be useful in enhancer identification. Both Creyghton *et al*. (2010) and Rada-Iglesias *et al*. (2011) described enhancer regions with nucleosomes exhibiting both H3K27ac and H3K4me1 modifications as being active, with enhancers where H3K4me1 alone is present being denoted as being inactive and 'poised' for activation. The latter form of enhancer are also enriched for the aforementioned repressive H3K27me3 mark (Rada-Iglesias *et al.* 2011) as illustrated in Figure 1.7. Bonn *et al*. (2012) expanded on this with the finding that H3K27me3 may also be used to distinguish enhancers with repressed activity rather than poised, with enhancers displaying this mark in mesodermal cells inactive, yet functioning in other cell types. There are further histone modifications that have been associated with enhancers, such as trimethylation of lysine 79 of histone 3 (H3K79me3) (Bonn *et al.* 2012), acetylation of lysine 64 of histone 3 (H3K64ac) and acetylation of lysine 122 of histone 3 (H3K122ac) (Pradeepa *et al.* 2016), however these have not been broadly adopted in the identification of putative enhancers. Yan *et al*. (2018) knocked out expression of the histone methyltransferase Myeloid/Lymphoid or Mixed-Lineage Leukaemia 3/4 (*MLL3/4*), observing that this led to a decrease in H3K4me1 deposition at enhancers and reduction in chromatin interactions. In addition, loss of *MLL3/4* also caused reduction in Mediator interaction, indicating that H3K4me1 acts in the recruitment of Mediator and therefore chromatin topological organisation (Yan *et al.* 2018). It would therefore be expected that this mechanism would result in developmental anomaly if perturbed. However, Rickels *et al*. (2017) knocked out expression of the *Drosophila* ortholog of *MLL3/4*, *Trr*, observing no profound effect on *Drosophila* development. Furthermore, in this study, the catalytic function of MLL3/4 was reduced in mouse embryonic stem cells (mESC), which caused a decrease in H3K4me1 deposition, but did not impair cellular proliferation. Each of

these studies emphasise the need for further elucidation of the extent and roles of histone modifications in enhancer function (Li *et al.* 2016).

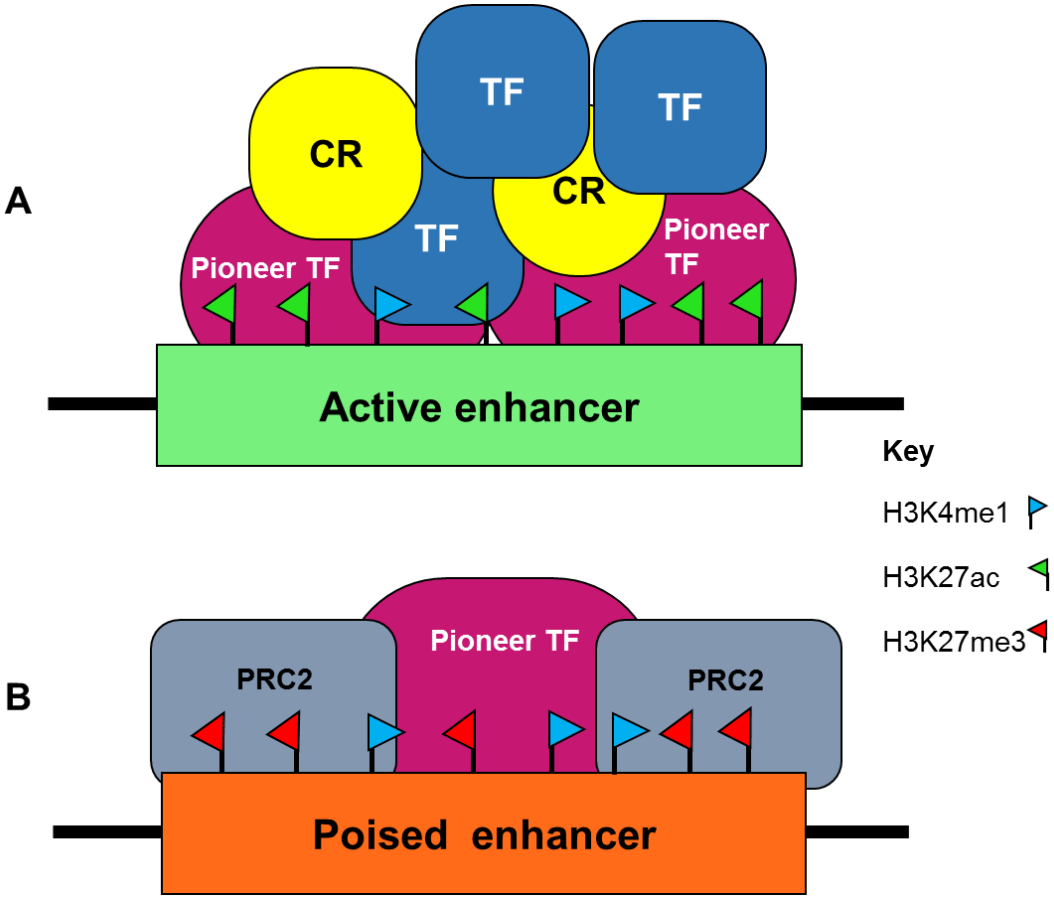


Figure 1.7: Chromatin state and interactions with regulatory proteins varies between active and poised enhancers. Active enhancers (A) display both H3K4me1 and H3K27ac histone modifications, with open chromatin structure that interacts with pioneer transcription factors (TF), chromatin remodelling proteins (CR) and coactivator TF. This is in contrast to poised enhancers which display H3K4me1 and the repressive H3K27me3 histone which is mediated by the Polycomb repressive complex (PRC2). This leads to chromatin structure that is able to interact with pioneer factors, yet the absence of cell-specific activator transcription factors prevents enhancer activity.

The ability to categorise enhancers on the basis of activity has had important ramifications in the understandings of the mechanisms involved in enhancer function. It is widely accepted that poised enhancers enable rapid induction of gene expression through a reduction in the number of steps required in order to induce initiation of gene transcription. Pioneer factors are capable of binding to poised enhancers which ensures that enhancer activity is triggered faster and more efficiently when the abundance of protein cofactors changes as the context of a cell transitions to specific conditions required for enhancer function (Figure 1.7) (Lagha *et al.* 2012; Levine *et al.* 2014). Poised enhancers are activated through interaction with chromatin remodelling proteins such as p300/CBP, in addition to cell lineage-

specific transcription factors, leading to a switch from H3K27me3 to H3K27ac with the loss of PRC2 repressive complex and the onset of chromatin looping, culminating in the enhancer and promoter being brought together; and the switch to active enhancer state (Plank and Dean 2014; Rada-Iglesias *et al.* 2011; Smith and Shilatifard 2014). Creyghton *et al*. (2010) suggest that enhancer poising is crucial in cell lineage determination, with Rada-Iglesias *et al*. (2011) and Bogdanović *et al*. (2012) reinforcing this with the notion that poised enhancers are marked at early stages of development in anticipation of activation in later stages of differentiation. However, many enhancers that are active do not transition from a poised to active state with changes in cellular differentiation state (Rada-Iglesias *et al.* 2011; Spitz and Furlong 2012).

1.3.3.3 Temporospatial specificity of enhancer function

As outlined previously, the most crucial feature of an enhancer is temporospatial specificity in function. The activity of most enhancers is delimited to a specific cellular lineage, which further reinforces cellular identity and the control of cell lineage (De Laat and Duboule 2013; Kieffer-Kwon *et al.* 2013). This attribute is particularly important, and utilised during development, with a significant consensus across the field that enhancers are critical determinants of cell lineage-specific patterns of gene expression and cell behaviour (Heinz *et al.* 2015; Kieffer-Kwon *et al.* 2013). This is exemplified by the work of Ernst *et al*. (2011) who found that enhancers are prevalent regulators of genes with developmental and tissue-specific roles; but not housekeeping genes. Moreover, housekeeping genes typically have fewer enhancers Osterwalder *et al.* (2018). The concept of housekeeping function is also posited by Zabidi *et al*. (2015), with the notion that developmental enhancers function in a highly specific cell type, whereas housekeeping enhancers function in multiple cell types. High-throughput methodology has enabled enhancer activity to be compared between cellular differentiation states. For example, Nord *et al*. (2013) identified 90,000 putative enhancers and studied their activation during development through the profiling of H3K27ac modification. Lineage-determining transcription factors interact with enhancers, with the abundance of these factors changing with developmental state, therefore moderating gene transcriptional output in a cell-specific manner (Kieffer-Kwon *et al.* 2013). Huang *et al*. (2016) demonstrated modified enhancer activity in erythropoietic stem cells as differentiation proceeded, and that differences in enhancer utilisation between cell stages was accounted for by combinatorial

interactions with coactivator proteins, rather than lineage-determining TF; the abundance of which was similar between lineages.

The stability and timing of enhancer-promoter interaction is a factor in enabling temporospatial specificity of function. Jin *et al*. (2013) observed enhancer-promoter interaction that had been forged prior to the presence of signals required in order to induce enhancer function in response to external stimuli. This is supported by Ghavi-Helm *et al*. (2014) who described looping interactions between enhancers and promoters as occurring in a stable manner. Paused Pol II has also been shown to reside concurrently with these interactions; albeit not with the replete PIC machinery. This suggests that Pol II stabilises these interactions and acts as a final major step in the switch from inactive to active gene transcription (Core *et al.* 2012; Ghavi-Helm *et al.* 2014; Lagha *et al.* 2012). The work of Fukaya *et al*. (2016) contrasts this, with the suggestion that enhancer-promoter interactions occur in a dynamic manner with interaction that occurs in a more transient manner for rapid transcriptional bursting. Pre-assembly of enhancer-promoter interaction is more frequently associated with tissue-specific enhancers that are rapidly activated in response to signalling triggered by cell-signalling effectors, such as in response to environmental stimuli (De Laat and Duboule 2013). Ostuni *et al*. (2013) found that poised enhancers can become activated in response to extracellular events. In addition, they described a further class of latent enhancers that were indistinguishable from surrounding non-functional chromatin on the basis of H3K4me1 or TF binding in unstimulated cells, but were activated in the presence of cognate cytokines; with the acquisition of enhancer marks, pioneer factor binding, enhancer activity and consequential gene transcription (Ostuni *et al.* 2013). This suggests that latent enhancer activation is a compromise between specificity in response to cognate stimuli and the timing involved in this. These enhancers have an increased threshold for activation given the lack of enhancer attributes that contribute to priming enhancer function, however, resultant activity is specific to stimuli and not inappropriately induced as could occur with pre-existing enhancer-promoter interaction. *De novo* enhancer-promoter interactions can be instigated by specific cellular context and are associated with more refined patterns of enhancer function; such as occurs with cell lineage-specific enhancers (De Laat and Duboule 2013). Rubin *et al*. (2017) examined enhancer-promoter interaction in cellular differentiation, describing two classes of such interaction. ‘Stable’ interactions existed prior to the commencement of differentiation with presence of cohesin as a facilitator of the interaction and no change in H3K27ac. ‘Gained’ interactions

mirrored poised enhancers with lack of cohesin, increase in H3K27ac, and increase in interaction strength upon differentiation. It remains to be seen whether future research will perpetuate this classification, however each of the aforementioned studies highlights the importance of cellular context as a determinant of enhancer purpose and activity.

Whilst temporospatial specificity is a fundamental feature that underpins enhancer function, there may be overlap in the activities of multiple enhancers for a cognate gene. Firstly, shadow enhancers are seemingly redundant *cis*-acting regulatory elements (Spitz and Furlong 2012). These forms of enhancer are located more distally from the target gene than other enhancers with highly similar identical patterns of TF binding and enhancer activity (Smith and Shilatifard 2014). However, shadow enhancers do not necessarily exhibit weaker function. Fukaya *et al*. (2016) observed greater induction of gene transcription facilitated by a shadow enhancer than a primary enhancer located proximal to the gene. This discrepancy in transcriptional output was due to increased bursting of transcription directed by the shadow enhancer. Cannavò *et al*. (2016) investigated the prevalence of shadow enhancers in *Drosophila* development, finding that shadow enhancers occur more commonly than previously envisaged and there may be multiple shadow enhancers for an enhancer, each of which are highly conserved. Redundancy in shadow enhancer function ensures robust gene expression (Lagha *et al.* 2012). For example, perturbation of the sequence and therefore function of one enhancer would be compensated for by the activity of a shadow enhancer being utilised in the same temporospatial context, preserving transcriptional output. The effects of this on topological organisation are unclear (Spitz and Furlong 2012). Redundancy between enhancers may be limited and only occur in a specific temporospatial window if enhancers display greater activity in divergent cell types (Cannavò *et al.* 2016). Osterwalder *et al*. (2018) used CRISPR-Cas9 based genome editing techniques in order to delete ten enhancers that have been proven to drive robust expression of genes involved in limb development. Knockout of these genes in mice causes limb malformation, and it was therefore expected that enhancer knockout would recapitulate these phenotypes. However, this loss-of-function approach did not cause limb associated phenotype for any enhancer that was deleted, thereby suggesting that redundancy between multiple enhancers for a gene.

1.3.3.4 Enhancers may be arranged within super-enhancers

Grouping of enhancer activity can also occur within super-enhancers. Whyte *et al*. (2013) first described super-enhancers as regions encompassing multiple enhancers and associated epigenetic modifications, with binding of cell lineage-determining transcription factors and enriched interaction with the Mediator protein complex. The function of multiple enhancers is therefore assimilated into stronger transcriptional output than from a typical single enhancer (Heinz *et al.* 2015; Huang *et al.* 2016). Nevertheless, the concept of super-enhancers is controversial. Hay *et al*. (2016) investigated the α-globin putative super-enhancer, with examination of multiple enhancer elements. They found that that there was variable function between enhancer elements within this region and that this resulted in additive function of individual elements as opposed to collaborative function. Organisation of sub-TADs is a further confusing aspect of this, as the organisation of a gene and its cognate enhancers into a sub-TAD grouping may confer super-enhancer identity. The recent work of Gong *et al*. (2018) illustrated that super-enhancers are insulated by strong TAD boundaries. This suggests that super-enhancers are closely regulated in order to ensure fidelity in function and that the potent transcription that arises from their function remains specific to the target gene. Further research is required in order to understand the veracity of super-enhancers and their function.

1.3.3.5 Enhancers may themselves be transcribed

The complexity in enhancer function is compounded by the fact that they may be transcribed when active. The seminal studies of de Santa *et al*. (2010) and Kim *et al*. (2010) described the occupation, and activity of Pol II at enhancer regions, catalysing the transcription of non-coding RNAs; which the latter group denoted as being enhancer RNA (eRNA). Bidirectional transcription of eRNA is more common than unidirectional transcription and whilst resultant transcripts can be polyadenylated, they typically have a short half-life and exhibit nascent nuclear localisation (Natoli and Andrau 2012; Rahman *et al.* 2016). A pertinent question in the field is whether eRNA is simply a consequence of Pol II presence at an enhancer, or an intrinsic aspect of enhancer mediated regulation of gene transcription. There is a growing consensus that eRNA function is important in transcriptional regulation, with several mechanisms as to how this is mediated, however this has not been extensively elucidated (Mikhaylichenko *et al.* 2018).

Enhancers are not transcribed in the absence of interaction with the target gene promoter, eRNA production is therefore a consequence of the interaction between enhancer and promoter (Kim *et al.* 2010; Zhu *et al.* 2013b). In addition, enhancer activity has been correlated with enhancer transcription (Mikhaylichenko *et al.* 2018). Arner *et al*. (2015) used cap analysis of gene expression (CAGE) methodology in order to examine the timing of transcription from promoters and enhancers, observing that active enhancers are transcribed before transcription of the cognate target gene. Mousavi *et al*. (2013) demonstrated that depletion of eRNA through short interfering RNA (siRNA) based silencing reduced chromatin accessibility and the loading of Pol II at cognate promoter region. This is echoed by Lam *et al*. (2013) who demonstrated that transcription of eRNA can be repressed by regulatory protein, leading to down-regulation of gene expression. This therefore suggests that eRNA have a role in the assembly of the PIC at the promoter and subsequent initiation of transcription. The effect of eRNA on histone modification was also echoed by the work of Kaikkonen *et al*. (2013) who described transcription from latent enhancers as leading to methylation of the histones of the enhancer. This group suggest that histone modification provides a form of epigenetic memory with the marking of enhancers for further activity. Enhancer RNAs have also been demonstrated to function in enhancer-promoter looping interaction. Li *et al*. (2013) observed that eRNA are capable of interaction with cohesin, and function in stabilising interaction between enhancer and promoter. These findings are contradicted by Schaukowitch *et al*. (2014) and Rahman *et al*. (2016) who found that eRNA are not required for enhancer-promoter interaction, with the former study demonstrating that reduction of eRNA abundance did not negatively impact on enhancer-promoter looping interaction. A further confounding factor in understanding eRNA is that function may occur during the transcriptional bursting process. Schaukowitch *et al*. (2014) described eRNA interaction with transcriptional pausing mediator NELF, which facilitated the release of Pol II and the commencement of the elongation phase of transcription. Each of these studies highlight the need for further interrogation of the role of eRNA in enhancer function. Moreover, recent work by Mikhaylichenko *et al*. (2018) has demonstrated the capacity for enhancers to act as promoters, and vice versa; further reiterating the complexity involved in attempting to understand enhancer function.

1.3.3.6 Enhancers are evolutionarily conserved

A further major trait associated with enhancers is a high degree of sequence conservation between evolutionary disparate species (Pennacchio *et al.* 2006). Genetic sequences with function, such as enhancers, are observed to be maintained in the genome through generations as a result of selective pressure, whilst non-functional regions are subject to genetic drift (Visel *et al.* 2007a). The conservation of TFBS is a critical aspect of enhancer conservation, ensuring that regulatory interactions can be maintained; and therefore an enhancer may still elicit function over evolutionary time (Taher *et al.* 2011). Loss of enhancer function may also contribute to evolution. Kvon *et al*. (2016) demonstrated that the process of loss of limbs in snake evolution could have been facilitated through inactivating mutations within the ZRS enhancer of limb development-related gene sonic hedgehog (*SHH*). CRISPR-Cas9 based replacement of murine enhancer genomic sequence with corresponding snake genomic sequence led to limb truncation in homozygous transgenic mice. This phenotype was rescued with the introduction of 17bp of evolutionarily conserved sequence present in the lizard but not the snake genome into the snake ZRS sequence in transgenic mice (Kvon *et al*. 2016). Dickel *et al*. (2018) illustrated that 'ultraconservation' over hundreds of base pairs is required for the function of some developmental enhancers, with developmental abnormality resulting from sequence aberration.

The disruption of DNA regions, such as TADs can have pathological consequences as mentioned previously (Lupiáñez *et al.* 2015). This is also true with aberration of enhancer DNA sequences and resultant function, which has been implicated in many diseases (Murakawa *et al.* 2016). Pathology may result from loss of enhancer-regulatory protein interaction if the regulatory protein has been mutated in a manner that diminishes capacity to recognise target DNA sequences or co-activator proteins (Smith and Shilatifard 2014). Much of the study of the role of enhancers in disease has however focused on changes in the DNA sequences of enhancers. For example, Maurano *et al*. (2012) examined the prevalence of disease associated single nucleotide polymorphisms (SNP) within putative regulatory elements predicted through the identification of DNase I hypersensitive regions. This study demonstrated that SNP within enhancers may be linked to phenotypes observed due to pathological changes in target gene expression. Subsequent improvements in whole genome methodology has enabled characterisation of the roles of many more enhancer mutations in pathology

(Murakawa *et al.* 2016). The manipulation of enhancer sequences and restoration of function through genome editing therefore represents a powerful future potential therapeutic approach in the amelioration of profound diseases. Moreover, greater elucidation of enhancer regions within the genome will facilitate a much greater understanding of many pathologies and the mechanisms that underpin them.

1.4 Transcriptional regulation of Ccn2/CCN2

Given that CCN2 is expressed in many tissues concurrently, the transcription of *CCN2* must be tightly controlled in order to ensure cell and tissue specificity in function. Many of the aforementioned studies that found changes in *Ccn2* expression after treatment with cellular signalling factors did not explore the mechanisms underlying this regulation. Characterisation of the *Ccn2/CCN2* promoter region has been an essential aspect of understanding regulation of transcription and therefore protein production. Many studies have used reporter assays in order to examine the ability of cellular signalling factors to regulate the transcriptional output from the human *CCN2* promoter (Eguchi *et al.* 2017).

Several transcription factors and cellular signalling mechanisms have been described as modulating *Ccn2*/*CCN2* expression via the promoter region, and are outlined in Figure 1.8. For example, the transcription factor specificity factor 1 (Sp1) binds to the *CCN2* promoter close to the coding sequence, within a region between -86b and +17bp relative to the TSS (Holmes *et al.* 2003). Furthermore, increased expression of *CCN2* was mediated by Sp1 in fibroblasts derived from scleroderma patients. The first study to scrutinise TFBS within the *CCN2* promoter was that of Grotendorst *et al*. (1996). This group used a fragment spanning -823bp to +74bp of the human *CCN2* promoter sequence to drive the expression of luciferase reporter gene. Treatment with TGF-β led to approximately 30 times greater reporter expression compared to untreated control in NIH3T3 fibroblast immortalised cell line. This study identified a TGF-β response element (Tβ-RE) located approximately 150bp upstream of the *CCN2* TSS. Response from this element was also found in HCS-2/8 cells (Eguchi *et al.* 2001). This region was later demonstrated to be required for basal *CCN2* promoter activity and became known as the basal control element (BCE) (Holmes *et al.* 2001). Endothelin-1 (ET-1) based regulation of *CCN2* expression is also dependent on utilisation of BCE-1, in addition to Mitogen-associated protein kinase kinase (MEK)/extracellular signal-regulated kinases (ERK) signalling (Shi-Wen *et al.* 2004).

Subsequent studies have discovered further regulatory motifs that mediate TGF-β control of *CCN2* expression. Abraham *et al*. (2000) found that TGF-β and tumour necrosis factor-α (TNFα) regulate *CCN2* transcription though motifs between -244bp and -166bp upstream of the *CCN2* TSS. (Figure 1.8) These proteins had antagonistic effects, with TGF-β stimulating transcription whereas TNF-α ablated this. Moreover, this TGF-β response element was further upstream of the previously described Tβ-RE/BCE (Figure 1.8). This work was further refined by Holmes *et al*. (2001) who showed that it was a Smad binding motif that mediated TGF-β response in this region. Smad transcription factors function downstream of TGF-β and BMP, all of which are members of the TGF-β superfamily of proteins (Massagué *et al.* 2005). Activation of TGF-β in turn leads to receptor triggered phosphorylation of Smad2 and Smad3, which form complexes with Smad4 and translocate to the nucleus where they can interact with co-activator proteins such as p300 and ultimately TFBS in coordinating gene transcription (Morikawa *et al.* 2013). Holmes *et al*. (2001) found that Smad 3 and 4 synergistically increase *CCN2* transcriptional output. Furthermore, in this study, TGF-β response was dependent on the function of Smad3. Smad 2 and 3 are also involved in TGF-β response in osteoblasts (Arnott *et al.* 2008). TGF-β regulation of *CCN2* expression in fibroblasts is also mediated through protein kinase C and Ras/MEK/ERK, and repressed by c-Jun N-terminal kinase (JNK) signalling (Leask *et al.* 2003). ERK functions downstream of both SRC and actin filament-associated protein (AFAP1) in TGF-β based *CCN2* promoter response in osteoblasts (Arnott *et al.* 2008; Cho *et al.* 2015). Leask *et al*. (2003) also described a transcription enhancer factor (TEF/TEAD) binding motif that was required for TGF-β regulation within the *CCN2* promoter in fibroblastic cells, thereby implicating interaction with the Hippo signalling pathway. However, further work by van Beek *et al*. (2006) revealed that Ets-1 is involved in TGF-β based *CCN2* promoter regulation and interacts at the TEF/TEAD motif. In addition, Ets-1 was found to function in the presence of Smad3 and was required in TGF-β response. A further transcription factor, friend leukaemia integration-1 (Fli-1) was also shown to interact alongside Ets-1. Ets-1 has also been shown to mediate TGF-β based *CCN2* promoter regulation in osteoblasts (Geisinger *et al.* 2012).

Utilisation of consensus sequences within the *CCN2* promoter is dependent on cell type. For example in the study from Grotendorst *et al*. (1996), TGF-β response occurred in fibroblastic and smooth muscle cells, but not epithelial cells. This is

supported by Eguchi *et al*. (2001) who compared basal *CCN2* promoter activity between HCS-2/8 chondrocytic cell line and HeLa cell line, finding that a 110 fragment located from -202bp to -88bp upstream of the *CCN2* TSS was more active in the chondrocyte derived cells. These studies therefore demonstrate that promoter activity is dependent on cellular context, and that chondrocyte specific response elements reside in this region. This work was expanded on with the discovery of the 'transcription enhancer dominant in chondrocytes' (TRENDIC) element located -202bp to -180bp upstream of the *CCN2* TSS (Eguchi *et al.* 2002). Utilisation of this element was found to be HCS-2/8 chondrocyte cell line-specific, further reinforcing the concept of *CCN2* transcriptional regulation in a cell type specific manner. Further study of TRENDIC demonstrated that it is bound by MMP3 which results in increased transcriptional output directed from the *CCN2* promoter (Eguchi *et al.* 2008).

Further chondrocyte specific regulation of *CCN2* transcription is mediated through a SOX9 TFBS. Firstly, Huang *et al*. (2010) described SOX9 and T-cell factor-lymphoid enhancer binding factor (TCF-LEF) consensus sites in the regulatory region surrounding *Ccn2* in the murine genome, and their utilisation in an *in vitro* model of endochondral ossification with regulatory sequence function linked to luciferase reporter gene expression. SOX9 and TCF-LEF recognise consensus site (A/T)(A/T) ACAA(A/T)G and inverted repeat sequence respectively, thereby facilitating antagonism between the two signalling mechanisms (Huang *et al.* 2010; Oh *et al.* 2010). Huang *et al*. (2010) found TCF-LEF sites at -443bp, -542bp, -1963bp and -3941bp upstream of the *Ccn2* TSS (Figure 1.8). In this study, β-catenin signalling mediated by TCF-LEF increased *CCN2* promoter activity and the expression of *CCN2* through the -443bp motif. SOX9 was observed to bind to this motif in a monomeric form that repressed gene expression. It was therefore postulated that TCF-LEF and SOX9 compete in order to bind to this motif and control *Ccn2* transcription. In this study, SOX9 was predicted to repress the expression of *Ccn2* in the primitive stages of mesenchymal condensation and chondrogenesis, with *CCN2* activation by TCF-LEF in latter stages of endochondral ossification as TCF-LEF expression and function becomes dominant over SOX9. The concept of antagonism between SOX9 and β-catenin is logical given the differences in the points at which they function in skeletogenesis and this has been previously proven (Akiyama *et al.* 2004). A puzzling aspect of the results obtained in this study is that a sequence from approximately -4kb to the TSS drove β-galactosidase (*LacZ*) reporter gene *in vivo* in intervertebral disc (IVD) chondrocytes

but not growth plate cartilage during embryonic development. This confounds previous work and their own study that found *Ccn2* expression in several populations of growth plate chondrocytes (Aoyama *et al.* 2009; Ivkovic *et al.* 2003; Kubota and Takigawa 2011).

Oh *et al*. (2016) also interrogated the capacity of SOX9 to regulate *CCN2* expression, however their data suggests that SOX9 promotes *Ccn2* transcription. This group used an *ex vivo* model with conditional knockout of *Sox9* expression in costal chondrocytes, finding that removal of *Sox9* function led to a decrease in *CCN2* expression. ChIP methodology revealed that SOX9 binds near to the *CCN2* TSS in the rat genome. A reductionist approach was subsequently used with luciferase reporter plasmids with different sized *Ccn2* promoter constructs in order to find the sequence that responds to SOX9. Mutation of a SOX9 motif -70bp to -64bp upstream of the TSS in the rat genome led to a decrease in reporter gene expression, moreover this site was bound by SOX9. This contradicted the findings of Huang *et al*. (2010) by firstly not discovering the previously described motif, and secondly with the observation monomeric SOX9 motif was utilised in increasing *CCN2* expression. Furthermore, Oh *et al*. (2016) also found that *Ccn2* expression within the growth plate of *Sox9* conditional knockout mice was reduced compared to control mice. The discrepancies between these studies highlight the importance of chondrocyte differentiation state in the context of *Ccn2* gene transcription.

Hiyama *et al*. (2018) also found that Wnt/β-catenin signalling can regulate the expression of *CCN2* in rat IVD chondrocytes. This study utilised a *CCN2* promoter reporter construct, with an increase in β-catenin function leading to decreased reporter expression. However, it is unclear how large a fragment of the *CCN2* promoter was used to drive reporter expression and there were no attempts to further refine the sequence within the construct that mediates the regulatory mechanism.

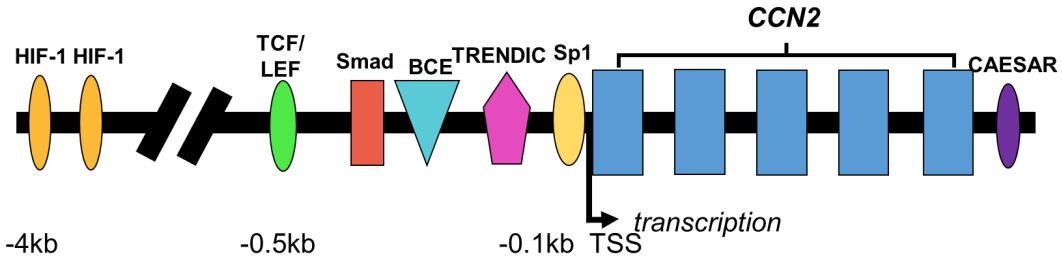


Figure 1.8: Basic schematic representation of transcription factor binding sites within the promoter region for *Ccn2/CCN2* in both mouse and human sequences. Many signalling mechanisms regulate the expression of CCN2. TGF-β is a potent regulator of CCN2 with control mediated through the Smad, BCE and Sp1 elements. Binding site utilisation is dependent on transcription factor abundance, and therefore cell type-specific.

Further studies have examined broader regulatory region encompassing kilobases of sequence upstream of *CCN2*. (Chiou *et al.* (2006) used a zebrafish model with green fluorescent (GFP) reporter gene expression under the control of a region spanning approximately 3kb upstream of *CCN2*. Treatment with growth hormone (GH) in addition to insulin-like growth factor 1 and 2 (IGF-1 and 2 respectively) caused an increase in promoter activity. Higgins *et al*. (2004) found hypoxia inducible factor-1 (HIF-1) binding motifs at approximately -3740bp and -1550bp upstream of *Ccn2* within the murine genome. These motifs were utilised in hypoxia response leading to an increase in *Ccn2* expression. This contrasts the findings of Tran *et al*. (2013) who assessed hypoxia response in nucleolus pulposus cells, finding that hypoxia caused a decrease in *Ccn2* expression and that HIF-1 reduced *Ccn2* expression, but this effect was not mediated by the HIF-1 response element motifs.

The 3' untranslated region (3'UTR) of *CCN2* also has regulatory function. This was first discovered by Kubota *et al*. (1999) who found that a segment of the 3'UTR repressed the expression of a luciferase reporter gene, and later named it as *cis*-acting element of structure-anchored repression (CAESAR) (Kubota *et al.* 2000). This therefore demonstrated that production of CCN2 is also regulated at the posttranscriptional level. Kondo *et al*. (2006) expanded on this by demonstrating hypoxic conditions in the culture of HCS-2/8 led to interaction between the CAESAR element in CCN2 mRNA and protein that caused greater stability of the mRNA and therefore increased translational output of CCN2.

As part of a project exploring genes expressed in the brain, Gong *et al*. (2003) created a transgenic mouse (Tg(Ctgf-EGFP)FX156Gsat) containing a bacterial

artificial chromosome (BAC) construct containing a GFP gene in the location of the endogenous translation initiation codon for *Ccn2.* Therefore, GFP reporter gene expression would be a surrogate for endogenous *Ccn2* expression driven by the surrounding regulatory sequence. This construct (RP24-96J1) is approximately 160kb in size and constitutes the region from chr10:24,258,010-24,418,466 in the mm9 mouse genome, which corresponds to approximately 60kb upstream of the TSS to approximately 100kb downstream of *Ccn2*. The resultant transgenic mouse was characterised at one month of age by Hall-Glenn and Lyons (2011) who described reporter gene expression within the cartilage of the intervertebral disc, costal cartilage, femoral growth plate and articular cartilage in addition to the cranial sutures and vasculature within the lungs, coronary arteries, and kidneys. Further examination of the function of this transgenic construct within the vasculature revealed reporter expression within arterial endothelium (Hall-Glenn *et al*. 2012). However reporter gene expression from the BAC construct does not completely recapitulate the endogenous expression of *Ccn2* within multiple tissues, suggesting that further regulatory elements outside that region are responsible for the complete expression profile of the gene. Moreover, the broad region that this BAC encompasses has not been further examined in attempts to refine understanding of *Ccn2* transcriptional regulation. Fundamentally, as of yet no enhancer related *cis*-regulatory module element has been described as a regulator of *Ccn2* expression. This is surprising given the range of cell types that the gene is expressed in, and the differential expression of the gene during development and beyond.

1.5 Aims

The lack of understanding on the capacity of *cis*-acting regulatory modules to regulate transcription of the *Ccn2* gene paved the way for the current study. The hypothesis of the current project was that enhancer regions would be present in the intergenic region upstream of murine *Ccn2*, and that combinatorial function of these *cis*-acting regions would drive temporospatial gene expression accounting for the endogenous pattern of *Ccn2* expression. Given the significance of *Ccn2* expression during cartilage and skeletogenesis, examination of enhancers governing *Ccn2* within skeletal tissues was given priority over other tissues.

Therefore the aims of the project were to:

1. ***Identify putative enhancer regions of Ccn2*** *in an in silico approach utilising publicly available datasets concerning enhancer associated chromatin characteristics in embryonic limb tissue.*

2. ***Assess the capacity of enhancers to drive reporter gene expression*** *in vivo, primarily during embryonic development, in addition to in vitro in cell lines*

3. ***Determine transcription factor binding motifs that mediate and regulate enhancer activity****, through the identification of TFBS in silico and validation of TF-DNA interaction in vitro using Electrophoretic Mobility Shift Assay*

2. Materials and Methods

2.1 In silico prediction of enhancer regions

The *Ccn2/CCN2* gene is still systematically known as *Ctgf*/*CTGF* in all *in silico* resources used.

The University of California Santa Cruz (UCSC) Genome Browser (https://genome.ucsc.edu) (Kent *et al.* 2002) available as part of the Encyclopaedia of DNA Elements (ENCODE) was used to visualise genomic loci of interest within the murine (*Mus musculus*) genome. The Genome Browser predominantly serves as a map of genomic sequence with capacity to integrate many annotations and datasets, these 'tracks', include gene function and structure, phylogenetic conservation and epigenetic modification of sequence within samples from a variety of tissues and cells at many time-points.

The mm9 (National Center for Biotechnology Information (NCBI) 37/mm9) assembly of the Genome Browser was primarily used. This genome build contains more annotations and a greater variety of datasets, than the newer murine mm10 (GRCm38/mm10) build, however tracks from mm10 and human hg18 (NCBI36/hg18) and human hg19 (GRCh37/hg19) assemblies were also used. The Genome Browser has a 'lift-over' tool which enables sequence coordinates to be converted between genome assemblies, therefore enabling equivalent sequences to be used regardless of source assembly build. The 'view DNA' function within the browser was used to gather genomic sequences for regions of interest in FASTA format.

A key aspect of the Genome Browser is the ability to add user generated custom tracks to the Browser, enabling tailored annotation of areas of interest. Custom tracks containing genomic coordinates for regions of interest were created through the conversion of basic text files (.txt) to browser extensible data (BED) (.bed) file format in order for upload onto the Genome Browser. For example, a custom track; 'regions of interest' with description 'CTGF/CCN2 enhancers' was created with each of the potential enhancer region assigned a colour, as demonstrated in Figure 2.1.

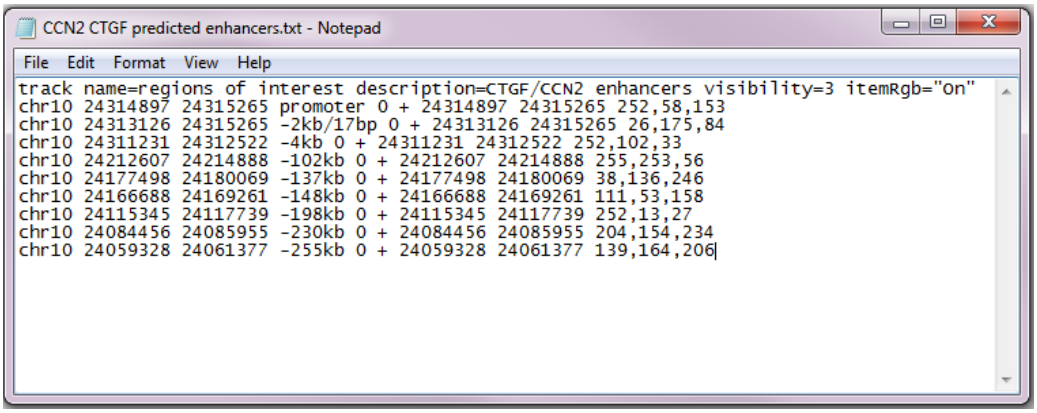


Figure 2.1: Creation of UCSC Genome Brower custom tracks. Basic text files are firstly created with entry of track name and description, region of interest coordinates, name, and orientation along with colour assignment for track points. These files are converted to BED format to enable upload onto UCSC Genome Browser.

The Genome Browser encompasses many datasets available from the ENCODE project that can be assimilated into the Browser. Firstly, sequence conservation information was obtained from the 30-Way Multiz Alignment resource (Blanchette *et al.* 2004). This track set enables comparative analysis of vertebrate DNA sequence with alignment of genomic sequence between species. Sequences from varying vertebrate clades were used, predominantly; mouse, rat, human, opossum, platypus, chicken, lizard and fugu. Tracks regarding DNase I hypersensitivity within forelimb bud, hind limb bud and mesodermal tissue at E11.5, in addition to lung fibroblast cells and NIH3T3 immortalised fibroblastic cells were used from ENCODE/University of Washington. ChIP-Seq tracks examining histone protein modification linked to transcriptional regulatory function were also incorporated into the Genome Browser session. Datasets from E14.5 limb tissue for the promoter specific H3K4me3, in addition to enhancer associated H3K4me1 and H3K27ac modifications were procured from ENCODE/Ludwig Institute for Cancer Research (LICR).

Further publicly available ChIP-Seq datasets regarding specific cell lineage related transcription factors were obtained from the NCBI's Gene Expression Omnibus (GEO) DataSets repository (https://www.ncbi.nlm.nih.gov/gds) (Barrett *et al.* 2012). Datasets were uploaded onto the UCSC Genome Browser in the BigWig or BED formats and incorporated into custom browser sessions. The aforementioned lift-over tool was used to convert datasets between genome builds. Datasets used can be found in Appendices Chapter 5.1.

2.2 Commonly used reagents and protocols

2.2.1 Nucleic acid manipulation

Mouse genomic sequences only were used for all experiments.

Invitrogen™ UltraPure™ diethyl pyrocarbonate (DEPC) -Treated Water (ThermoFisher Scientific, Waltham, Massachusetts, 750024) was used as the default nuclease free water in molecular techniques and in the suspension and dilution of nucleic acids.

All polymerase chain reaction (PCR) primer oligonucleotides were procured from Eurofins Genomics (Ebersberg, Germany). These oligonucleotides were unmodified with salt free purification. Lyophilised oligonucleotides were re-suspended to 100µM concentration as standard.

Deoxyribonucletotide triphosphates (dNTPs) (Bioline, London, United Kingdom, BIO-39025) were used in PCR methodology. Stock solutions were created containing 10µM each of deoxyadenosine triphosphate (dATP), deoxycytidine triphosphate (dCTP), deoxyguanosine triphosphate (dGTP), deoxythymidine triphosphate (dTTP) diluted with nuclease free water.

2.2.1.1 Primer design

The online tool 'Primer Basic Local Alignment Search Tool (Primer-BLAST)', available from NCBI at https://www.ncbi.nlm.nih.gov/tools/primer-blast/ was used to find primers suitable for generation of amplicons from genomic sequences. This tool was also used to find any potential non-specific amplicon that could be generated during PCR using the 'Primer Pair Specificity Checking Parameters' based on the *Mus musculus* genome. Alternately, primers were manually selected at the ends of regions of interest. Sequences of approximately 20bp were selected with guanosine/cytosine (GC) content of 40 to 60%, melting temperature (T_m) of approximately 60°C with G or C at the 3' end of the primer in order to promote specific binding. The T_m for primers in a pair were no more than 5°C different from one another. Primers designed manually were also examined in the Primer-BLAST tool in order to assess suitability.

Where restriction enzyme recognition sites were added to the 5' of primers, additional nucleotides required for restriction endonuclease recognition were added in line with the New England BioLabs 'cleavage close to the end' resource, available at: https://international.neb.com/tools-and-resources/usage-guidelines/cleavage-close-to-the-end-of-dna-fragments.

2.2.1.2 PCR

Purified mouse genomic DNA (Promega, Madison, Wisconsin, United States, G3091) was used as template in all non-genotyping PCR reactions. All PCR reactions were incubated in T100™ PCR Thermal Cycler (Bio-Rad, Hercules, California, United States) or SimpliAmp™ Thermal Cycler (ThermoFisher).

Where reaction efficiency was low with weak product formation, or where multiple products had formed, gradient PCR runs were conducted whereby replicate reactions were annealed at a range of temperatures in order to find optimum annealing temperature. Touchdown PCR was also used for difficult amplicons, this typically consisted of using an initial annealing temperature 5°C higher than the T_m of the primers before being reduced by 1°C per cycle for the first 10 cycles of the reaction before further cycling with annealing temperature predicted to be optimum for primer pair. Where optimum annealing temperature for a primer pair was within 1°C of the polymerase extension temperature, two step cycling approach was used whereby the annealing step was integrated into a longer extension step.

GoTaq® G2 Flexi DNA Polymerase

For the amplification of DNA amplicons of 2kb or less, GoTaq® G2 Flexi DNA Polymerase (Promega, M7805) was used as standard DNA polymerase unless otherwise stated.

Typical 25µL reaction set up consisted of:

Component	*Volume*	*Final concentration*
5X Colourless GoTaq® Flexi Buffer [1]	5µL	1X
$MgCl_2$ solution (25mM)	2µL	2mM
dNTP mix (10µM each dNTP)	0.5µL	200µM
Forward primer (10µM)	0.5µL	200µM
Reverse primer (10µM)	0.5µL	200µM
DNA template (100ng)	X[2]	4ng/µL
GoTaq® Flexi polymerase (5u/µL)	0.125µL	0.025U/µL
Nuclease free water	to 25µL	

Standard cycling parameters consisted of:

Stage	*Temperature*	*Duration*
Initial denaturation	94°C	3 minutes
30 cycles:		
Denaturation	94°C	30 seconds
Annealing	Y[3]	30 seconds
Extension	72°C	1 minute per kb product
Final extension	72°C	5 minutes
Hold	4°C	indefinite

Phire™ Hot Start II DNA Polymerase

For amplicons greater than 2kb in length, the Phire Hot Start II DNA Polymerase (ThermoFisher, F122S) was used, unless stated otherwise.

[1] For end-point PCR the 5X Green GoTaq® Flexi Buffer was used in place of the 5X Colourless GoTaq® Flexi Buffer.

X[2] Volume of DNA template was dependent on concentration of DNA and adjusted accordingly for a total of 100ng per reaction.

Y[3] Standard annealing temperature for each amplicon was typically 5°C lower than the lowest T_m in each primer pair.

Typical 20µL reaction set up consisted of:

Component	*Volume*	*Final concentration*
5X Phire Reaction Buffer	4µL	1X
dNTP mix (10µM each dNTP)	0.4µL	200µM
Forward primer (10µM)	1µL	500nM
Reverse primer (10µM)	1µL	500nM
DNA template	X[4]	5ng/µL
Phire™ Hot Start II DNA polymerase	0.4µL	Y[5]
Nuclease free water	to 20µL	

Standard cycling parameters consisted of:

Stage	*Temperature*	*Duration*
Initial denaturation	98°C	30 seconds
30 cycles:		
Denaturation	98°C	5 seconds
Annealing	Z[6]	5 seconds
Extension	72°C	15 seconds per kb product
Final Extension	72°C	1 minute
Hold	4°C	indefinite

Platinum™ *Pfx* Polymerase

Platinum™ Pfx Polymerase (ThermoFisher Scientific, 11708) was also used in the generation of longer PCR amplicon.

X[4] Volume of DNA template was dependent on concentration of DNA and adjusted accordingly for a total of 100ng per reaction.

Y[5] No details from the supplier regarding the concentration of the enzyme.

Z[6] Annealing temperatures used were typically 3°C higher than the lower primer T_m of a pair

Typical 25µL reaction consisted of:

Component	*Volume*	*Final concentration*
10x *Pfx* amplification buffer	5 µL	2x
dNTP(10µM)	0.75 µL	0.3µM
$MgSO_4$ (50mM)	0.5µL	1mM
Forward primer (10µM)	0.75µL	0.3µM
Reverse primer (10µM)	0.75µL	0.3µM
DNA template (100ng)	µL	4ng/µL
Pfx polymerase (1.25U/µL)	0.2µL	0.5U
Nuclease free water	to 25µL	

Standard cycling parameters consisted of:

Stage	*Temperature*	*Duration*
Initial denaturation	94°C	3 minutes
30 cycles:		
Denaturation	94°C	15 seconds
Annealing	X^7	30 seconds
Extension	68°C	1 minute per kb product
Final Extension	68°C	1 minute
Hold	4°C	indefinite

2.2.2 Agarose gel electrophoresis

Molecular grade agarose powder was procured from Sigma-Aldrich (St Louis, Missouri, United States, A9539) and Bioline (BIO-41025). Tris-acetic acid-ethylenediaminetetraacetic acid (EDTA) (TAE) buffer system was used for all gels. 50x stock buffer was prepared using the following recipe: 242g Tris base (Merck Millipore, Burlington, Massachusetts, United States, 648310), 57.1mL glacial acetic acid (Sigma Aldrich, A6283), 18.6g EDTA disodium salt dehydrate (VWR, Radnor, Pennsylvania, United States, 20296.291) with distilled water to total volume of one litre. This was diluted to 1X working solution using distilled water. Gels were typically cast using 1% w/v agarose in 1X TAE buffer. Gels were stained using

X^7 Annealing temperature was typically 5°C lower than that of the lowest primer T_m of a pair

Ethidium Bromide (EtBr) solution (Promega, H5041) at final concentration of 100pg/µL, or SYBR Safe DNA Gel Stain (ThermoFisher Scientific, S33102). The latter is supplied as a 10,000X concentrate, which was used at 1 in 25,000 dilution. Sub-Cell® and Mini-Sub® cell agarose gel electrophoresis systems (Bio-Rad) were used for all gels. 6x Purple Gel Loading Dye (New England BioLabs, B7024S) was used to prepare samples for electrophoresis at a final concentration of 1X. This was omitted where samples already contained gel loading dye as a PCR buffer constituent. Quick-Load® Purple 2-Log DNA ladder (1 kb Plus DNA ladder), Quick-Load Purple 1kb Plus DNA Ladder and λ-HindIII Digest (New England BioLabs, Ipswich, Massachusetts, United States, N3200S, N0550S and N3012S) were typically loaded at 500ng per well and used for determination of fragment size and concentration. Gels were run at voltage of 10V/cm.

Gels were visualised using Gel Doc XR system (Bio-Rad) or Genegenius system (Syngene, Bangalore, India).

2.2.3 Gel extraction

Gels were visualised using a UV transluminator. In order to avoid mutagenesis, gels were left in the casting tray and the wavelength of the transluminator was set at 365nm with 80% emittance rate. Fragments of interest were excised quickly using a clean scalpel.

The QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany, 28706) was used to purify DNA from gel fragments, in accordance with the manufacturer's instructions. All centrifugation steps were carried out at room temperature and consisted of one minute spins at 16,000*g*. Briefly, fragments were weighed before addition of Buffer QG of volume 3x that of the gel fragment mass, after which samples were incubated at 55°C in a waterbath. Samples were removed at two minute intervals and vortexed in order to aid gel disintegration. When no visible trace of gel was observed, molecular grade isopropanol (Sigma Aldrich, I9516) was added to the sample at a volume equal to that of the gel fragment mass. Samples were then applied to QIAquick column before centrifugation, after which flow-through was discarded. 750µL Buffer PE was then added to the column before centrifugation. Flow through was discarded and column was centrifuged again in order to remove residual buffer. Columns were transferred to clean microcentrifuge tubes before

30µL nuclease free water was added to the centre of each column with incubation at room temperature for five minutes prior to centrifugation.

For fragments greater than 2kb, nuclease free water was heated to 55°C prior to addition to the column. Where replicate samples were purified, the product of one column was used to elute the DNA from the second column with two stages of incubation prior to centrifugation.

Yield of gel extraction product DNA was assessed using agarose gel electrophoresis with comparison of sample band intensity to ladder fragments of known mass (Chapter 2.2.2)

2.2.4 Column PCR product purification

PCR products were purified using the Monarch® PCR and DNA Cleanup Kit (New England BioLabs, T1030S), in accordance with the manufacturer's guidelines. PCR product was diluted in Binding Buffer at ratio of 2:1 buffer to DNA for fragments larger than 2kb and 5:1 buffer to DNA for amplicon smaller than 2kb. Centrifugation steps consisted of one minute spins at 16,000*g*. Diluted samples were applied to a Monarch column prior to centrifugation, after which flow-through was discarded. 200µL Wash buffer was then added to the column before centrifugation. Flow-through was discarded before wash step was repeated. Columns were then transferred to clean microcentrifuge tubes prior to addition of 30µL nuclease free water to the centre of each column and incubation at room temperature for five minutes. As with the gel extraction protocol, multiple reactions were pooled at the elution stage.

Concentration of purified products was assessed through spectrophotometry using Nanodrop™ 2000 (ThermoFisher Scientific).

2.2.5 Restriction digests

Where multiple restriction enzymes were used within a reaction the 'NEBcloner Digestion: Restriction Enzyme Selection'; available at https://nebcloner.neb.com/#!/redigest, was used to find the most efficient incubation temperature and buffer conditions for digestion.

DNA restriction digests were typically double digests and consisted of 4µg DNA, 1µL restriction enzyme one (typically ten units), 1 µL restriction enzyme two (typically ten units), 2.5µL of 10X buffer (NEBuffer 1.1, 2.1, 3.1 or Cutsmart® Buffer (New England Biolabs, B7200S)).

Reactions were incubated at 37°C, with the exception of those containing ApaI. This restriction enzyme was used at room temperature as its optimal temperature is 25°C. Where ApaI was used in double digests, reactions were incubated at room temperature overnight with ApaI before addition of second enzyme and increase in temperature to 37°C for further digestion for two hours.

2.2.6 Bacterial culture

Lennox L (LB) broth and agar (Invitrogen®, 12780-052 and 22700-025 respectively) were prepared in accordance with manufacturer's instructions and used as standard media for the propagation of *E. coli* bacteria.

2.3 Cloning and plasmid DNA extraction

2.3.1 Vectors

'Serial Cloner' (Serial Basics) freeware software was used to visualise and manipulate plasmid sequences and sites of interest *in silico* for each vector generated. Cloning was carried out prior to the adoption of *Ccn2* as the official gene symbol, therefore *CTGF* is used to denote the gene.

2.3.1.1 pCR™8/GW/TOPO®TA

Each full enhancer region, aside from -230kb and -255kb, was cloned into the pCR™8/GW/TOPO®TA vector by Dr Ian Li (ThermoFisher Scientific, K2500). This is a basic plasmid used in the maintenance and propagation of sequences of interest. The plasmid contains the prerequisite sequences required for maintenance in bacteria and is used to create insert containing entry vectors for the Gateway™ (GW) cloning system (Invitrogen, ThermoFisher Scientific, Hartley *et al*. (2000)), with map of sequences outlined in Figure 2.2.

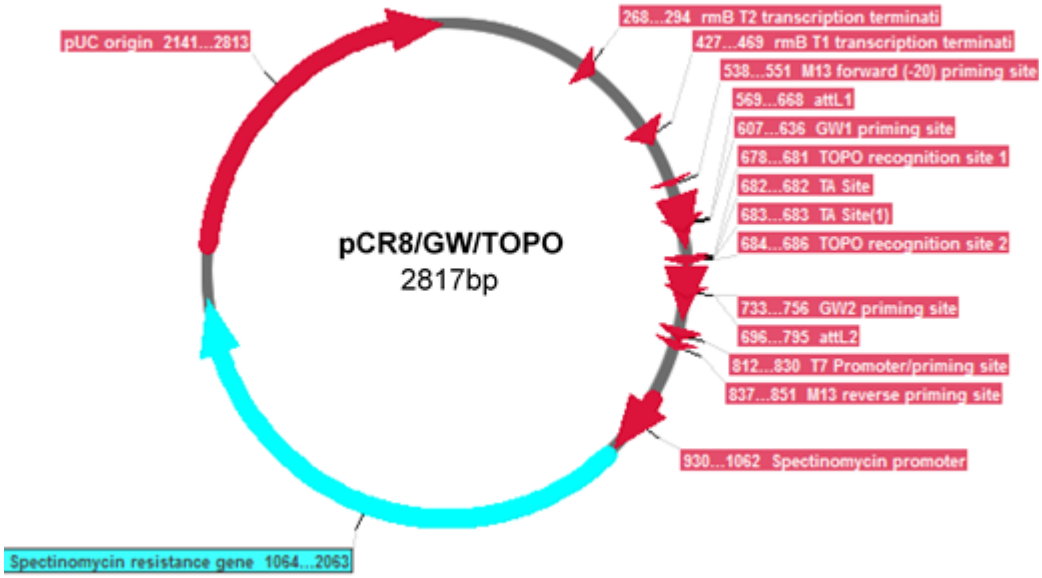


Figure 2.2: pCR™8/GW/TOPO®TA. This is an entry vector for gateway cloning. Insert is placed in vector using TA cloning. The insert is cloned into the vector using the TA sites. The vector contains a spectinomycin resistance gene for bacterial selection, in addition to a pUC origin for plasmid replication within bacteria.

2.3.1.2 β-galactosidase reporter vectors

Each enhancer region was cloned into reporter plasmids based on the Heat shock protein 68- β galactosidase (Hsp68-*lacZ*) vector, which was a gift from Dr Nadav Ahituv (Pennacchio *et al.* 2006). Based on the work of Kothary *et al*. (1989), this plasmid contains *lacZ* derived from *E. coli* which encodes the β-galactosidase metabolic enzyme and is commonly used as a reporter in transgenic animal models. Expression of the gene is mediated by an Hsp68 promoter region which is incapable of directing tissue specific expression. Therefore, an enhancer sequence placed upstream of both the promoter and *lacZ* regulates reporter gene expression. Two strategies were used for this; either the use of TOPO entry vectors combined with the Hsp68-*LacZ*-GW destination vector, or sticky end cloning of inserts into an Hsp68-*LacZ* vector which has been modified to remove the GW sites.

The Hsp68-*lacZ*-Gateway plasmid is a destination vector in the Invitrogen™ Gateway™ system which is based on sites of recombination within entry and destination vectors (Figure 2.3). The -4kb, -102kb, -137kb, -148kb, -198kb and -255kb Hsp68LacZGW vectors were created by Dr Ian Li, with plasmid maps in the Appendices (Chapter 5.2.1)

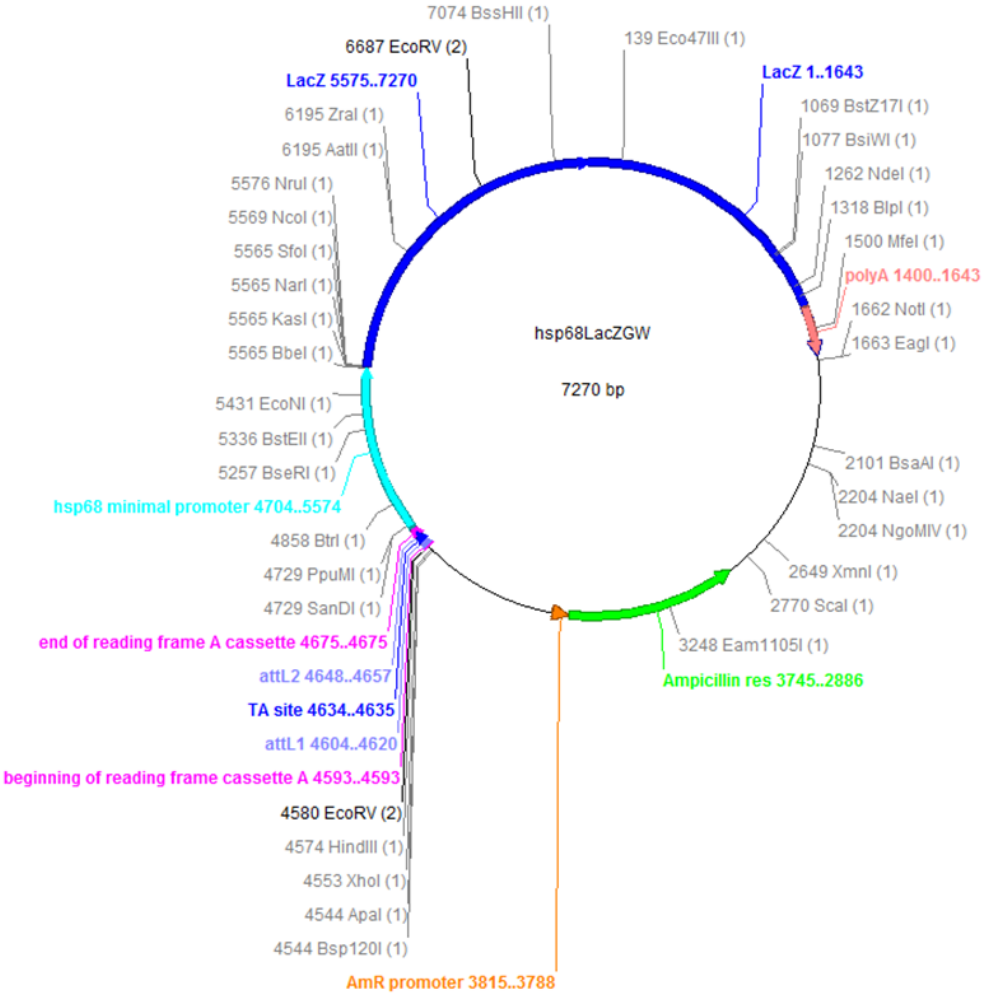


Figure 2.3:Hsp68LacZGW basic plasmid map. This gateway destination vector contains recombination sites (*att*) enabling sequence of interest to be inserted upstream of the Hsp68 minimal promoter and *LacZ* coding sequence (dark blue). The vector also contains ampicillin resistance gene for bacterial selection (map courtesy of Dr Ian Li)

A modified version of this vector was used for the creation of the -137kbshortHsp68LacZ and -230Hsp68LacZ vectors (Figure 2.4) with plasmid maps detailed in Appendices Chapter 5.2.2. This plasmid does not have the sites required for Gateway® recombination, therefore inserts were sub-cloned into the vector using sticky-end restriction cloning.

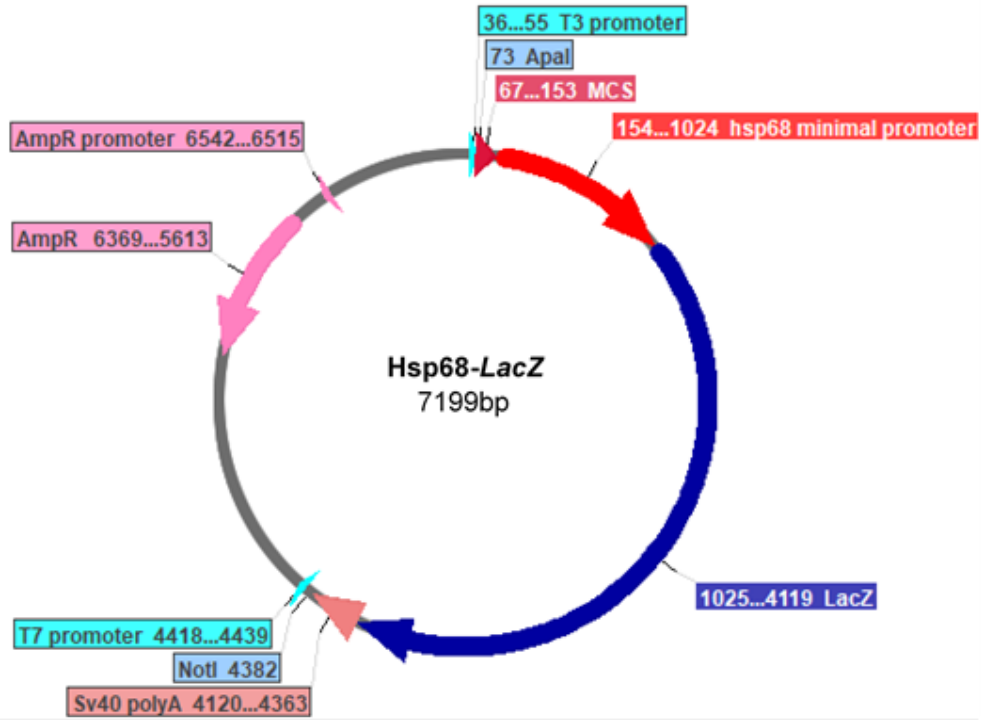


Figure 2.4: Hsp68-*LacZ* plasmid map. This vectors contains the Hsp68 minimal promoter region and *LacZ*, but does not contain sites required for Gateway® recombination. Inserts are cloned into the vector upstream of the Hsp68 minimal promoter, therefore enabling an enhancer to regulate the expression of *LacZ*. Selection in bacteria is based on ampicillin resistance.

Inserts were cloned into the plasmid using sites within the multiple cloning site upstream of the Hsp68 promoter. Primers were designed for inserts with a restriction enzyme recognition site and required additional base pairs for efficient cutting.

2.3.1.3 Luciferase reporter vectors

The pGL4.10 plasmid (Promega, E6651) was used as the basis for enhancer containing luciferase expression vectors (Figure 2.5). This plasmid contains firefly (*Photinus pyralis*) *luc2* reporter gene which is codon optimised for mammalian expression, in addition to a synthetic β-lactamase gene for ampicillin resistance based bacterial selection. The basic plasmid has no promoter region, but contains a multiple cloning site (MCS) for fragment insertion.

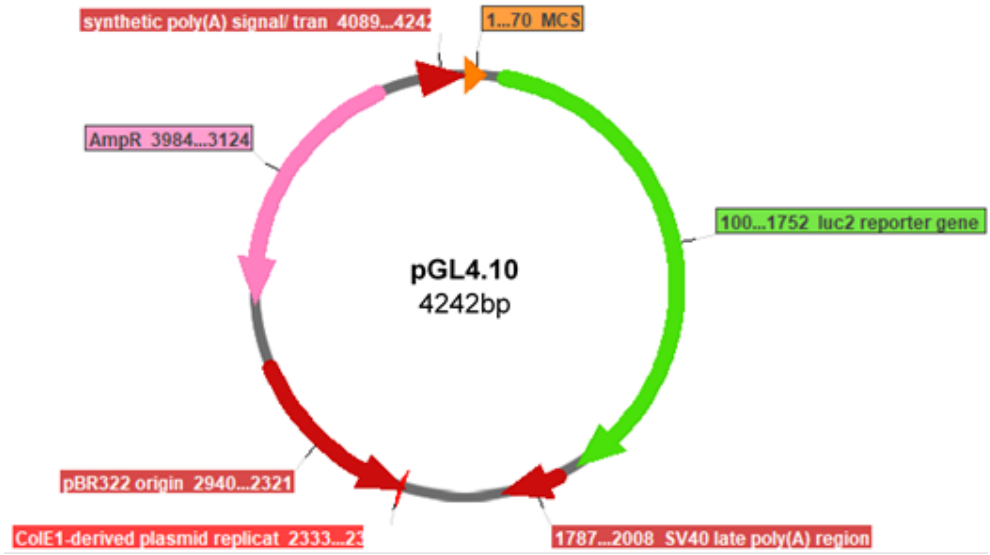


Figure 2.5: pGL4.10 luciferase reporter basic plasmid map. This reporter plasmid does not have an enhancer or promoter region, enabling the introduction of custom regulatory elements within the multiple cloning site (MCS) which are used to regulate the expression of a firefly luciferase gene (luc2).

The *Ccn2* promoter region spanning from -353bp to +17bp of the *Ccn2* coding sequence was used to drive the expression of *luc2*. This promoter sequence was selected based on consensus from literature that has demonstrated that this region drives gene expression (Leask *et al.* 2003). This region was inserted upstream of *luc2* using KpnI restriction site at the 5' of the insert and HindIII site at the 3' end (Figure 2.5). The resultant plasmid became known as pSLF01 (plasmid map detailed in Appendices, Chapter 5.2.3). Each enhancer region of interest was inserted upstream of the promoter region using KpnI site at the 5' end of the insert, with either SacI or NheI site at 3' end of the insert. The -255kb enhancer was inserted using an XhoI site at the 3'. This allowed *luc2* expression to be regulated by the enhancer sequence via the *Ccn2* promoter. Each of the enhancer containing variants of pSLF01 is detailed in Appendices Chapter 5.2.3

2.3.2 Vector assembly

2.3.2.1 pCR™8/GW/TOPO®/TA

Creation of pCR™8/GW/TOPO enhancer vectors was carried out by Dr Ian Li (Frost *et al.* 2018). Insert DNA was amplified using PCR with the REDExtract-N-Amp polymerase (Sigma Aldrich, E3004). This polymerase has leaves 3' A overhangs which are required for integration of product into pCR™8/GW/TOPO® TA. Primers used are listed in Table 2.1

Enhancer	Coordinates (mm9)	Primer name	Sequence 5' to 3'	T_m (°C)	Product length (bp)
-4kb	chr10: 24311246-24311268	-4 CTGF Fw	CTCTACCCCAGACTACAGGTTCA	64.2	1295
	chr10: 24312517-24312541	-4 CTGF Rv	TGAAAGGGTGAAGACAGAAGACTAC	61.3	
-102kb	chr10:24212606-24212628	-102 CTGF Fw	ACCAGATCAGACACCGAGCAATA	60.6	2282
	chr10:24214866-24214888	-102 CTGF Rv	TGGTTAATGGCTCACGTGGATTC	60.6	
-137kb	chr10:24177497-24177520	-137 CTGF Fw	GAAGCGCAAGAAGGAAGACCAAAG	62.7	2575
	chr10:24180046-24180069	-137 CTGF Rv	CAGCTCCTTTGCCTTTGCACTGTA	62.7	
-148kb	chr10:24166687-24166708	-148 CTGF Fw	TTCTTGCAGATGTGCTGAGGTG	60.3	2574
	chr10:24115344-24115366	-148 CTGF RV	GAAAGATGGAGGGTTAGAGACAAG	61.0	
-198kb	chr10:24115344-24115366	-198 CTGF Fw	GGTCTTAGGCAAGCAAATCTCTG	60.6	2395
	chr10:24117717-24117739	-198 CTGF Rv	CATTGAAGAGTCCAAGAAGCAGG	60.6	
-255kb	chr10:24059327-24059350	CTGF -255 F	CAGAATGCCTGAGTGAGATAGAAG	61.0	2050
	chr10:24061349-24061377	CTGF -255 R	TGTTTGTGAATTATCTAGCAAGGGAAAAG	61.0	

Table 2.1: Primers for amplification of basic enhancer region sequences. These primers were used for the amplification of the full enhancer region sequences. Resultant amplicon were placed in pCR™8/GW/TOPO®/TA vector using TA cloning. These primers were also used for general amplification of the enhancers

Typical PCR reaction set up consisted of:

Component	*Volume*	*Final concentration*
2x REDExtract-N-Amp PCR reaction mix	10µL	1x
Forward primer (10µM)	0.8µL	0.4µM
Reverse primer (10µM)	0.8µL	0.4µM
DNA template (100ng)	X [8]	5ng/µL
Nuclease free water	to 20µL	

Typical cycling parameters:

Stage	*Temperature*	*Duration*
Initial denaturation	94°C	3 minutes
30 cycles:		
Denaturation	94°C	30 seconds
Annealing	Y[9]	30 seconds
Extension	72°C	1 minute per kb product
Final extension	72°C	10 minutes
Hold	4°C	indefinite

Specificity of PCR products was assessed using agarose gel electrophoresis.

TOPO® cloning reactions were set up using manufacturer's guidelines as follows: 2µL PCR product, 1µL salt solution, 2µL RNase/DNase free water and 1µL TOPO® vector. These reactions were incubated at 25°C for five minutes before being placed on ice prior to bacterial transformation. 2µL of these reactions was then added to a vial of One Shot™ TOP10 chemically competent *E. coli* before being incubated on ice for 30 minutes. Transformation reactions were then placed at 42°C for 30 seconds prior to incubation on ice for two minutes. 250µL super optimal broth with catabolite repression (SOC) media (Invitrogen, 15544-034) was then added to reactions before shaking incubation at 37°C for one hour. Reactions were then plated on LB agar plates containing 50µg/mL spectinomycin (Sigma Aldrich, S4014) before overnight incubation at 37°C.

X[8] Template volume varied depending on concentration of stock solution

Y[9] Annealing temperature of 5°C lower than the lowest primer T_m was used as standard.

2.3.2.2 Gateway™ recombination

For each of the regions listed in Table 2.1 *lacZ* reporter destination vectors were created using the Invitrogen™ Gateway Recombination™ Technology. Plasmid inserts were generated using the aforementioned pCR8/TOPO entry vector (Chapter 2.3.2.1). Each enhancer insert was therefore flanked by *attL* sites within the pCR8/TOPO vector. The Hsp68-*lacZ*-GW vector contains *attR* sites upstream of the Hsp68 promoter region. The LR Clonase™ II system (ThermoFisher) facilitates recombination between these sites, and therefore the incorporation of the enhancer insert upstream of the promoter region and *lacZ* reporter gene.

Briefly, 5µL of enhancer containing pCR8/TOPO entry vector DNA was mixed with 1µL Hsp68-lacZ-GW vector DNA, 2µL Tris-EDTA (TE) buffer (pH 8.0), 2µL LR Clonase™ II enzyme mix and incubated at 25°C for one hour. 1µL proteinase K solution was added to each reaction before incubation of samples at 37°C for ten minutes. One Shot™ chemically competent *E. coli* were then transformed as before, with 100µg/mL ampicillin (Sigma Aldrich, A9518) based selection rather than spectinomycin.

2.3.2.3 Sticky-end cloning

Insert generation

For -137kbshortCTGFHsp68LacZ, -230kbCTGFHsp68lacZ and all luciferase based reporter vectors, enhancer regions were amplified using primers with restriction enzyme sites incorporated into the 5' end as detailed in Table 2.2.

The *Ccn2* promoter was first cloned into the pGL4.10 vector, with the resultant pSLF01 vector being used as a backbone for subsequent cloning of enhancer regions. For cloning into luciferase vectors, inserts were amplified with pre-existing pCR8/TOPO enhancer plasmids as template where available, using primers flanking the inserts. Primers amplifying from within vector backbone sequence with restriction enzyme sites added at the 5' end which were designed by Dr Ian Li. PCR product specificity was verified using agarose gel electrophoresis. DNA fragments were purified using either the QIAquick gel extraction kit or Monarch® PCR and DNA Cleanup Kit; as outlined in sections 2.2.3 and 2.2.4 respectively.

Region of interest	Destination vector	Primer name	Sequence 5' to 3'	T_m (°C)	Product length (bp)
-137kb short enhancer	Hsp68-LacZ	137s+ApaI_F	GAgggcccGCATTTCCAAAGGGGAGACCTGGA	74.6	1299
		137s+XhoI_R	GTACctcgagGTGAGCCCAGTGAATGTCCTTG	72.0	
-230kb enhancer	Hsp68-LacZ	-230+ApaI_F	GAgggcccGGGCAATTTTAACAAGGCTGAGTA	70.8	1441
		-230+HindIII_R	GCCaagcttTCTCAGGTTCTCAGTCAGTTCTTT	68.2	
Ccn2 promoter	pSLF01	CTGFPRXho_F	GCATCTCGAGGGCCCATGGTATTTGCCTCTTGAG	73.1	388
		CTGFPRHind_R	CGTAAGCTTCCGGCTCGCCAAAGAACTGA	69.5	
-137kb short enhancer	pSLF01	137Kpn_F	GTGGTACCGCATTTCCAAAGGGGAGACCTGGA	72.0	1298
		137Nhe_R	GACGCTAGCGTGAGCCCAGTGAATGTCCTTG	72.1	
-230kb enhancer	pSLF01	SF_CTGF230 kpnI F	GTGGTACCGGGCAATTTTAACAAGGCTGAGTA	68.2	1444
		SF_CTGF230 NheI_R	GACGCTAGCTCTCAGGTTCTCAGTCAGTTCTTT	69.5	
-255kb enhancer	pSLF01	SF_-255+NheIF	CTGGCTAGCCAGAATGCCTGAGTGAGATA	68.1	2047
		SF_-255+XhoIR	CGTGCTCGAGGTTTGTGAATTATCTAGCAAGG	68.2	
-4kb, -137kb full, -148kb, -198kb enhancers	pSLF01	prlL16KpnI	TGATGGTACCAGCTCGGGCCCCAAATAATG	69.5	1572, 2849, 2851, 2672
		prlL17NheI	TGACGCTAGCGGGATATCAGCTGGATGGCAAAT	70.9	
-102kb enhancer	pSLF01	prlL16KpnI	TGATGGTACCAGCTCGGGCCCCAAATAATG	69.5	2559
		prlL17SacI	CATGGAGCTCGGGATATCAGCTGGATGGCAAAT	70.7	

Table 2.2: Primers for sticky-end cloning of enhancer reporter vectors. Primers contained restriction enzyme recognition site and required extra bases on 5' end for sticky-end cloning into reporter gene plasmids. The prlL primers were designed by Dr Ian Li, with complementary sequence in the TOPO sequence flanking inserts. This enabled them to be suitable to amplify any insert that had been cloned into the pCR8/TOPO vector backbone

Ligation

Inserts were restriction digested before purification through agarose gel electrophoresis and gel extraction (Chapter 2.2).

Vector backbone and insert fragments were ligated using T4 DNA ligase (New England BioLabs, M0202). The molar ratio for ligation was calculated using the following equation:

$$\frac{\text{insert size (bp)}}{\text{vector size (bp)}} \text{x vector amount (ng)=insert amount (1:1)}$$

Reactions were set up using the manufacturer's guidelines. A ratio of 4:1 insert: vector was typically used. For example, in the ligation of pSLF01_137kbshort: vector backbone (pSLF01) size 4576bp, insert (-137kb short) size 1298bp, where vector amount is 50ng

$$\frac{1298}{4576} \text{x } 50 = 14.18$$

14.18ng of insert required for 1:1 ratio; 56.72ng for 4:1 insert to vector. Therefore, ligation reaction set up consisted of: 2µLT4 DNA Ligase Buffer, 50ng pSLF01 vector, 56.72ng -137kb short insert, 1µL T4 DNA Ligase and RNase/DNase free water to a total of 20µL. Reaction components were gently aspirated before brief centrifugation and incubation at 16°C for 16 hours. Reactions were heat inactivated with incubation at 65°C for ten minutes.

Transformation

DH5α (ThermoFisher Scientific, 18265017) and JM109 (Promega, L2001) strains of chemically competent *E. coli* were transformed with ligation products. 50µL aliquots of cells were allowed to thaw on ice before addition of 10ng ligation product with gentle agitation to mix. Samples were incubated on ice for 30 minutes. Heat-shock was carried out with incubation in a waterbath at 42°C for 20 seconds for DH5α and 50 seconds for JM109, after which tubes were returned to ice for two minutes. 450µL SOC media was then added to each reaction before shaking incubation at 225rpm and 37°C for one hour. Reactions were then plated at varying

densities, with 20µL, 50µL and 100µL spread on LB agar plates containing 100µg/mL ampicillin before plates were incubated at 37°C overnight.

Control transformation reactions were carried out using the pUC19 (ThermoFisher Scientific, SD0061) in order to ensure transformation efficiency, with 250pg of plasmid used to transform cells as above.

2.3.3 Colony PCR

Single colonies were sampled from bacterial transformation plates using pipette tip which was then placed in a microcentrifuge tube containing 100µL nuclease free water. This inoculum was vortexed before 1µL was used as template in PCR. Insert primers were used to generate amplicon with reaction parameters as used for insert preparation. PCR products were ran on agarose gels to confirm insert presence.

Insert positive colonies were propagated for plasmid DNA extraction (Chapter 2.3.4)

2.3.4 Plasmid DNA extraction

For propagation from glycerol stock, bacteria were firstly streaked across an LB agar plate containing suitable selection antibiotic before being incubated at 37°C overnight. Single colonies were isolated from agar plates from glycerol stock propagation and transformation plates using a pipette tip before addition to 3mL LB broth containing suitable selection antibiotic at appropriate concentration. Tubes were briefly vortexed before shaking (225rpm) incubation at 37°C overnight.

After this incubation period, cultures were briefly vortexed. 1.5mL of overnight culture was then added to a microcentrifuge tube before bacteria were pelleted through centrifugation at 9600*g* for three minutes. The supernatant was removed before the addition of a further 1.5mL of culture and repetition of centrifugation step and removal of supernatant.

2.3.4.1 PureYield™ Plasmid Miniprep System

Mini-preparation of plasmid DNA was conducted using PureYield™ Plasmid Miniprep System (Promega, A1223) in accordance with the manufacturer's

guidelines. All centrifugation steps were carried out at 16,200*g* at room temperature.

Overnight culture bacterial pellets were re-suspended in 600µL nuclease free water before addition of 100µL Cell Lysis Buffer and inversion of tubes six times. 350µL of cold Neutralisation Solution (4°C) was then added before tube was inverted six times. Samples were then centrifuged for three minutes. Supernatant was transferred to PureYield™ Minicolumn prior to centrifugation at for 15 seconds. Flow-through was discarded before 200µL Endotoxin Removal Wash was added to the column before centrifugation for 15 seconds. Flow-through was discarded before addition of 400µL Column Wash Buffer to column. Samples were then centrifuged for 30 seconds. Columns were then transferred to a clean microcentrifuge tube after which 30µL nuclease free water or TE buffer was added to centre of the column before incubation at room temperature for five minutes. Plasmid DNA was eluted with final centrifugation step of 15 seconds before storage at -20°C.

2.3.4.2 Crude mini-preparation of plasmid DNA

The protocol for crude mini-preps of plasmid DNA is based on that of the Qiagen Plasmid kit protocols, utilising buffers P1, P2 and P3.

Overnight bacterial culture pellet was re-suspended in 100µL buffer P1 (50mM Tris Cl pH 8.0, 10mM EDTA, 10µg/mL Ribonuclease (RNase) A, LyseBlue) (Qiagen), ensuring that no visible clumps of bacteria remained. 100µL buffer P2 (200mM sodium hydroxide (NaOH), 1% sodium dodecyl sulfate (SDS) (w/v)) (Qiagen) was then added before gentle inversion six times, ensuring suspension was homogenous and incubation for five minutes at room temperature. 100µL buffer P3 (3M potassium acetate pH 5.5) was then added before gentle inversion six times or until there was no visible LyseBlue blue dye in suspension. Samples were then incubated on ice for five minutes before centrifugation at 16,200*g* for ten minutes. 750µL of cold molecular grade absolute ethanol (Sigma Aldrich, 51976) (kept at -20°C) was then added to the resultant supernatant before gentle inversion twice, after which samples were placed at -20°C for ten minutes. After this period, samples were inverted twice before centrifugation at 16,200*g* for ten minutes. The supernatant was discarded before 1mL 70% molecular grade ethanol was added to sample and subsequently removed in order to wash plasmid DNA pellet. Pellets

were left to air dry at room temperature for approximately ten minutes and then re-suspended in 30µL TE buffer (10mM Tris Cl pH 8.0, 1mM EDTA) and left to reconstitute at room temperature for one hour or at 4°C overnight. Plasmid DNA was stored at -20°C.

2.3.4.3 Maxi-preparation of plasmid DNA

Larger scale plasmid DNA extraction was carried out using the Plasmid Maxi Kit (Qiagen, 12162). All centrifugation steps were carried out at 4°C.

Bacterial starter cultures were created from glycerol stocks as aforementioned, however these cultures were incubated for eight hours rather than overnight. 200µL of this starter culture was then transferred to 100mL LB broth containing appropriate selection antibiotic and incubated at 37°C overnight with shaking at 225rpm.

Bacterial cells were pelleted through centrifugation at 6000*g* for 15 minutes before re-suspension in 10mL Buffer P1. 10mL Buffer P2 was then added before inversion six times; ensuring homogenous suspension, and incubation at room temperature for five minutes. 10mL of cold (4°C) Buffer P3 was the added before further inversion six times and incubation on ice for 20 minutes. Samples were inverted once and then centrifuged at 20,000*g* for 30 minutes. The supernatant was removed prior to further centrifugation at 20,00*g* for 15 minutes. Resultant supernatant was then applied to QIAGEN-tip 500 which had been pre-equilibrated through application of 10mL Buffer QBT (750mM sodium chloride (NaCl) 50mM MOPS pH 7.0, 15% isopropanol (v/v), 0.15% Triton® X-100 (v/v)). Samples were allowed to pass through the column before washing with 30mL Buffer QC (1M NaCl, 50mM MOPS pH 7.0, 15% isopropanol (v/v)). This wash step was repeated before application of 15mL Buffer QF (1.25M NaCl, 50mM Tris-HCl pH 8.5, 15% isopropanol (v/v)). Eluate from this step was collected before addition of 10.5mL molecular grade isopropanol and gentle inversion. Samples were then centrifuged for 30 minutes before supernatant was discarded. 5mL 70% ethanol was then added to pellet before centrifugation at 15,000*g* for ten minutes. Supernatant was discarded and plasmid DNA pellets were left to air-dry for approximately ten minutes. Pellets were re-suspended in 500µL nuclease free water before being left at 4°C overnight.

2.3.5 Glycerol stocking

After verification of vector insert sequence presence and orientation using colony PCR and restriction digests, 500µL overnight bacterial culture was added to 500µL 50% glycerol (v/v with distilled water) (Sigma Aldrich, G5516). The resultant solution was gently inverted to ensure homogenous mixture before being snap-frozen on dry ice and stored at -80°C.

2.4 CRISPR-Cas9 based genome editing

CRISPR-Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats-Caspase 9) technology has galvanised the field of molecular biology in allowing targeted manipulation of specific genomic sequences. There are many emerging variants of CRISPR-Cas9 technology, however for the current project an approach of inducing double strand breaks (DSB) in the DNA sequence with wild-type *Streptococcus pyrogenes* Cas9 (SpCas9) was used. In this variant of the methodology, CRISPR-Cas9 function hinges on the association of guide RNA (gRNA) that recognises target sequences, and Cas9 protein that induces double strand breaks in genomic sequence. Single guide RNA (sgRNA) contain a protospacer sequence which is complementary to target genomic DNA sequence (Cong *et al.* 2013). This DNA sequence is found next to a protospacer adjacent motif (PAM) that is recognised by the complexed Cas9 (Nowak *et al.* 2016). The PAM for SpCas9 is -NGG. A *trans*-activating CRISPR RNA (tracrRNA) interacts with the sgRNA and facilitates Cas9 interaction and therefore formation of the gRNA-Cas9 complex required for the recognition and cleavage of target DNA (Jinek *et al.* 2012). Cas9 contains two endonuclease domains RuvC and HNH which cleave target DNA sequence to produce a DSB. For the purpose of this project, target loci were chosen in the proximity of the 5' and 3' boundary of enhancer regions. The enhancer sequences would therefore be excised and because there was no replacement of enhancer DNA with a donor sequence, DSB would be recognised and repaired using non-homologous end joining (NHEJ) mechanism. This would therefore lead to the loss of the enhancer region and joining of the sequence in the vicinity of the 5' boundary, to sequence in proximity of the 3' boundary of the enhancer.

The methodology used in the current project is based on that of Professor David Young at Newcastle University. Initially target loci and suitable gRNA were

identified *in silico*. Guide RNA and TRACR RNA were transcribed from a DNA template which was prepared through PCR from a vector template. The forward primer was specific for each gRNA used and mismatch in annealing between the primer and template allowed amplification of primer sequence and therefore gRNA template DNA. The transcript generated was referred to as gRNA and purified before *in vitro* validation of the ability of gRNA-TRACR-Cas9 to induce DSB in PCR product substrate. Finally, gRNA-TRACR and SpCas9 were microinjected into mouse embryos with the ambition of deleting *Ccn2* enhancers *in vivo*.

2.4.1 *In silico* design of guide RNA

The CHOPCHOP online tool (Labun *et al.* 2016; Montague *et al.* 2014), available at http://chopchop.cbu.uib.no/ was used to find suitable guide RNAs. This software utilises the mm10 build of the murine genome therefore loci of interest were converted from mm9 to mm10 using the lift-over function within the UCSC Genome Browser as detailed in Chapter 2.1.

Guides were selected to target loci within the vicinity of the 5' and 3' boundary of each enhancer (as illustrated in Figure 2.6). Two guides were chosen for each enhancer boundary loci. Coordinates based on the enhancer boundary +/-100bp were initially used to search for suitable guide RNA target sequences. Where this yielded guides with poor predicted efficiency the search region was expanded to +/-200bp from the enhancer boundary.

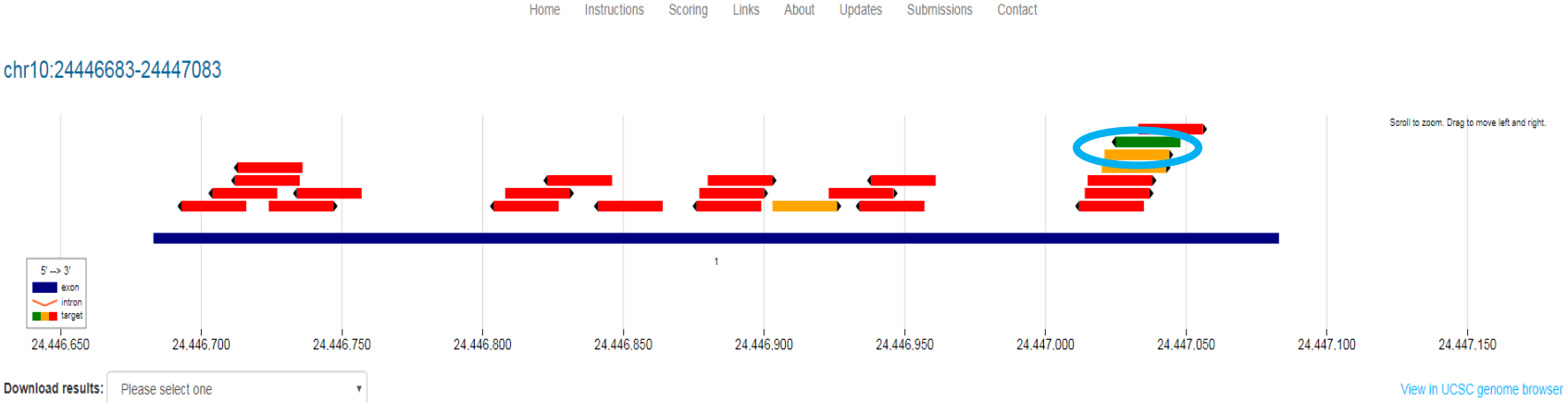
Figure 2.6: Schematic representation of strategy in selecting guide RNA for enhancer excision. Guides were chosen within 200bp (red dashed box) of the 5' and 3' boundary of each enhancer region of interest (black box).

Search parameters utilised consisted of: CRISPR/Cas9, with sgRNA length without PAM of 20 nucleotides (nt), 3' NGG as a PAM and off-targets were determined with up to three mismatches in the protospacer region (Hsu *et al.* 2013). Searches with efficiency scores based on the algorithms of (Xu *et al.* 2015) and (Moreno-Mateos *et al.* 2015) were used. The CHOPCHOP tool automatically ranks potential guides based on predicted off-targets and efficiency scoring. The highest ranked guides;

and therefore those predicted to have greatest efficiency, were chosen for each enhancer boundary.

For the excision of both the -148kb and -137kb enhancers simultaneously, guides were selected within loci more distal from the enhancer boundary. A region spanning 15kb and encompassing both of these enhancers was used in order to find a single guide upstream of the -148kb 5' boundary, with a second guide selected downstream of the -137kb 3' boundary.

For example, for the -148kb 5' boundary (chr10:24,446,883) a targeting region +/- 200bp (chr10: 24,446,683-24,447,083) was used as region of interest. This generated 24 potential sgRNA sequences of which the two best ranked were chosen (highlighted with blue circles Figures 2.7 and 2.8) (Appendices Chapter 5.3).



Ranking	Target sequence	Genomic location	Exon	Strand	GC (%)	Self-complementarity	Off-targets				Efficiency
							0	1	2	3	
1	ATGTCCCTAACCCTTGCTTATGG	chr10:24447026	1	-	45	1	0	0	0	0	0.45
2	TATGCCATAAGCAAGGGTTAGGG	chr10:24447022	1	+	40	0	0	0	0	1	0.44
3	TTATGCCATAAGCAAGGGTTAGG	chr10:24447021	1	+	40	0	0	0	0	2	0.50
4	CTACAATAAAATCATCTACTAGG	chr10:24446904	1	+	25	1	0	0	0	1	0.39
5	TGGTATATTTTAGGGAGTCGTGG	chr10:24446705	1	-	40	0	0	0	0	3	0.51
6	CCAGATAACACCTATTTACATGG	chr10:24446725	1	+	35	0	0	0	0	2	0.40
7	TATAATTATGCCATAAGCAAGGG	chr10:24447016	1	+	25	0	0	0	0	3	0.67
8	GGTGTTATCTGGTATATTTTAGG	chr10:24446714	1	-	30	0	0	0	0	3	0.28
9	CATGGTACACTCATTCCTCATGG	chr10:24446824	1	-	45	0	0	0	1	3	0.44
10	TTATATGCAAGTATTCCATGAGG	chr10:24446809	1	+	30	0	0	0	0	4	0.60
11	ATGGAATACTTGCATATAACTGG	chr10:24446805	1	-	30	0	0	0	0	4	0.35
12	AGAAGAGATTGCCTTTTGGAAGG	chr10:24446935	1	-	40	1	0	0	1	4	0.53

Figure 2.7: Screenshot of CHOPCHOP guide selection. This tool finds and ranks potential guide sequences within genomic regions of interest based on GC%, self-complementarity, and possible off-targets. Guides are ranked based on the efficiency with which they are predicted to mediate the cutting of the target sequence. The highest ranking sequences were used for guide synthesis (highlighted with blue circle)

The CHOPCHOP tool also allows assimilation of guide sequences into the UCSC Genome Browser in the form of a custom track (Figure 2.8). This therefore allows more accurate assessment of the proximity between potential guides and enhancer regions.

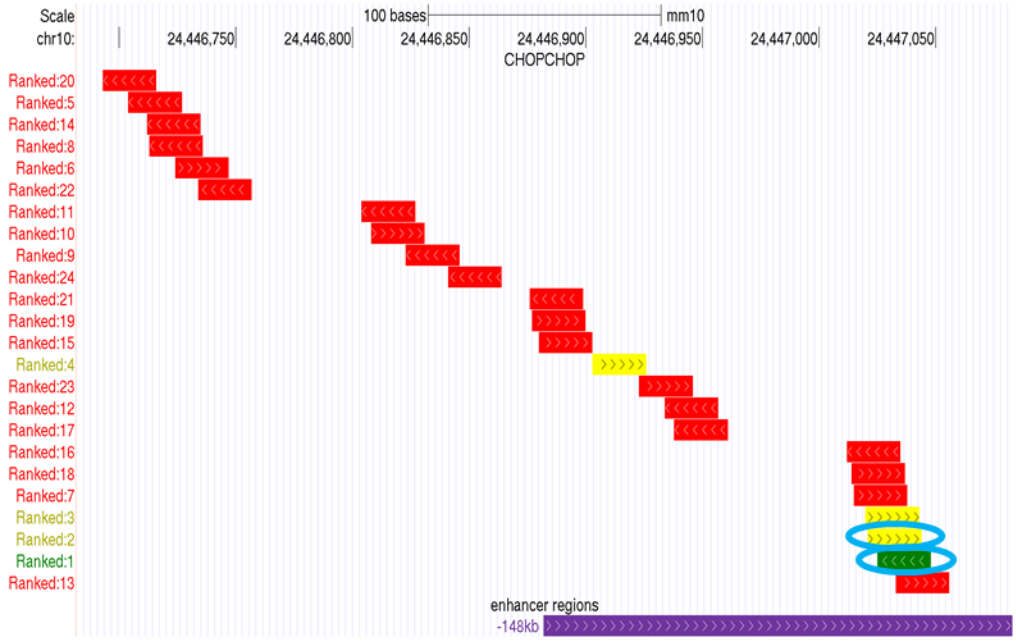


Figure 2.8: Incorporation of CHOPCHOP guides into custom UCSC Genome Browser session. CHOPCHOP predicted guides can be imported as a custom track into UCSC Genome Browser sessions, therefore allowing visualisation of guide position with respect to each enhancer region of interest.

2.4.2 Guide RNA synthesis and purification

The Tra2bgl in CRISPRV2 plasmid (a gift from Professor David Young, Newcastle University) was used as template in the generation of DNA precursor for guide RNA. This plasmid is a modified version of pX330-U6-Chimeric_BB-CBh-hSpCas9 (Hsu *et al.* 2013) containing a guide RNA sequence that has been cloned in upstream of a TRACR sequence. The basic pX330 plasmid is illustrated in Figure 2.9.

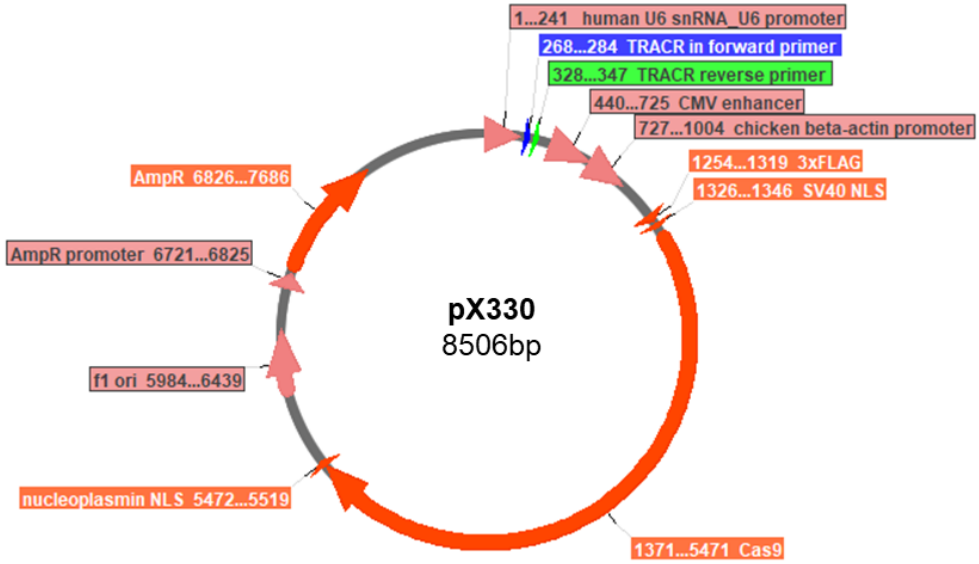


Figure 2.9: pX330 CRISPR guide Cas9 expression plasmid This plasmid contains a TRACR sequence that is required for CRISPR gRNA function (Hsu *et al.* 2013). In PCR amplification of guide template, the 3' end of the forward primer used (highlighted pale blue) and the reverse primer (green) both amplify the TRACR sequence. Whilst the plasmid contains sequence encoding Cas9, this was not used in the current approach.

2.4.2.1 Synthesis and purification of DNA template for guide RNA transcription

Guide sequences found from CHOPCHOP were incorporated into a long forward primer (Figure 2.10) specific for each guide which contained a T7 promoter for *in vitro* transcription, the guide sequence without the PAM and the 5' extremity of the TRACR motif sequence required for the secondary structure of the RNA

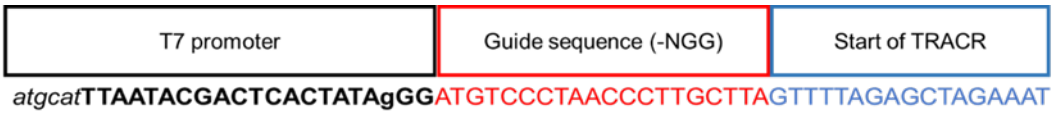


Figure 2.10: Schematic representation of a long forward primer used for amplification of gRNA template DNA. A T7 promoter was upstream of the guide sequence specific for each target site in addition to the 5' end of the TRACR sequence. The guide sequence of the primer would mismatch with the vector template DNA in PCR annealing, leading to amplification of the guide-specific sequence rather than vector DNA.

Each guide specific forward primer (listed in Table 2.3) was used in conjunction with a reverse primer that had sequence complementary to the 3' end of the TRACR sequence (highlighted green in Figure 2.9) with sequence: 5'-AAAAGCACCGACTCGGTGCC. The amplicon for each guide template was approximately 120bp long. GoTaq® Flexi G2 DNA polymerase was used for PCR, as outlined in Chapter 2.2.1.2. In brief, 20ng of plasmid DNA was used as template in a 50µL reaction. Touchdown PCR was used in order to reduce the effect of large primer Tm discrepancy on reaction efficiency.

Enhancer region	Guide name	Sequence
-137kb	-137_5'_1	*atgcat***TTAATACGACTCACTATAgGG**CTGGTATCCACAATGAGGTTGTTTTAGAGCTAGAAAT
	-137_5'_2	atgcat**TTAATACGACTCACTATAgGG**TGGTATCCACAATGAGGTTCGTTTTAGAGCTAGAAAT
	-137_3'_1	atgcat**TTAATACGACTCACTATAgGG**TATGGAGTTTTGCTGAACTGGTTTTAGAGCTAGAAAT
	-137_3'_2	atgcat**TTAATACGACTCACTATAGG**GTTGACTCTAAACTATTATGGTTTTAGAGCTAGAAAT
-148kb	-148_5'_1	*atgcat***TTAATACGACTCACTATAgGG**ATGTCCCTAACCCTTGCTTAGTTTTAGAGCTAGAAAT
	-148_5'_2	*atgcat***TTAATACGACTCACTATAgGG**TATGCCATAAGCAAGGGTTAGTTTTAGAGCTAGAAAT
	-148_3'_1	*atgcat***TTAATACGACTCACTATAgGG**CTCAGGACTACAATTGCATCGTTTTAGAGCTAGAAAT
	-148_3'_2	*atgcat***TTAATACGACTCACTATAgGG**CCGTCTTACAAAACATATAAGTTTTAGAGCTAGAAAT
-148kb/-137kb	-148/137_5	*atgcat***TTAATACGACTCACTATAgGG**TTGATAGTGGCCACAAGAGAGTTTTAGAGCTAGAAAT
	-148/137_3	*atgcat***TTAATACGACTCACTATAGG**GTGAATCCCAAATTGTGACTGTTTTAGAGCTAGAAAT
-230kb	-230_5'_1	*atgcat***TTAATACGACTCACTATAgGG**AGTTACCATAAGCCCACCAGGTTTTAGAGCTAGAAAT
	-230_5'_2	*atgcat***TTAATACGACTCACTATAgGG**TAAGGCTTTACATTAGCGAAGTTTTAGAGCTAGAAAT
	-230_3'_1	*atgcat***TTAATACGACTCACTATAgGG**AGCACAGATGATGTCGTTTGGTTTTAGAGCTAGAAAT
	-230_3'_2	*atgcat***TTAATACGACTCACTATAgGG**AAGAGCAATGATGTGCTATAGTTTTAGAGCTAGAAAT

Table 2.3: Forward primers used for the generation of gRNA DNA template. Each primer contained the T7 promoter (italic and bold), specific targeting sequence and 5' segment of the TRACR sequence.

Assessment of PCR efficiency and specificity was made using agarose gel electrophoresis (Chapter 2.2.2). 5µL of PCR product for each guide was ran on a 2% agarose gel in order to assess reaction specificity and efficiency. PCR product was purified using the Monarch® PCR and DNA Cleanup Kit (Chapter 2.2.4). Resultant product was used as template in the transcription of the guide RNA.

2.4.2.2 In vitro transcription and guide RNA purification

An RNase free environment was used throughout the preparation of gRNA.

Guide RNA was produced using the MEGAshortscript™ T7 Transcription Kit (ThermoFisher Scientific, AM1354) which utilises a T7 polymerase, in accordance with the manufacturer's guidelines. 150nM column purified PCR product was used as template for each reaction. Transcription reactions were prepared at room temperature as follows: 2µL T7 10x Reaction Buffer, 2µL 75mM T7 adenosine trisphosphate (ATP) solution, 2µL 75mM T7 cytidine triphosphate (CTP) solution, 2µL 75mM T7 guanosine triphosphate (GTP) solution, 2µL 75mM T7 uridine triphosphate (UTP) solution, 2µL T7 Enzyme Mix, with template DNA and nuclease free water to a final volume of 20µL. Reactions were incubated in a thermal cycler at 37°C for 4 hours, after which 1µL of Turbo DNase was added to each reaction before further incubation at 37°C for 15 minutes.

Resultant RNA was immediately purified or stored at -80°C prior to purification.

The pTRI-RNA 18s template supplied in the MEGAshortscript™ T7 Transcription Kit was used as a positive control in accordance with the manufacturer's guidelines.

Guide RNA was purified using the MEGAclear™ Transcription Clean-Up Kit (ThermoFisher Scientific, AM1908) in accordance with the manufacturer's guidelines. 80µL Elution solution was added to each transcription reaction product, after which samples were gently mixed before addition of 350µL Binding Solution Concentrate was added and samples were aspirated gently. 200µL absolute molecular grade ethanol was then added and samples were gently aspirated to mix. Samples were applied to the filter cartridge and centrifuged at 9600*g* for one minute. Flow-through was discarded before 500µL Wash solution was added to each column with centrifugation at 9600*g* for one minute. This step was repeated after flow-through was discarded. Columns were then centrifuged at 9600*g* for 30

seconds in order to remove any residual Wash Solution. 50µL embryo water (Sigma Aldrich, W1503) was then added to each column which were incubated at 70°C for ten minutes prior to final elution centrifugation step of 9600*g* for one minute. RNA concentration and purity was assessed using Nanodrop.

2.4.3 CRISPR cutting assay substrate preparation

2.4.3.1 Substrate primer design

In addition to guide prediction, the CHOPCHOP tool also enables the selection of primer pairs for *in vitro* assessment of sgRNA-Cas9 cutting efficacy. Search parameters can be changed to customise product length, primer length, primer T_m and minimum distance between primer and guide target site. Search parameters used were: 300bp to 1000bp product length, primer size of 18nt to 25nt with optimum of 22nt, primer T_m of 55°C to 70°C with an optimum of 65°C and minimum distance between guide cutting site and primer of 50bp. Primers were selected so that cutting of substrate fragment would not lead to fragments of around 100bp in length that could be confused with any gRNA band on gel. Potential primer pairs were entered into the Primer-Blast tool as outlined in Chapter 2.2, in order to confirm product specificity and predict primer self-complementarity.

A primer pair was selected for each enhancer boundary region. Where the guides for a boundary were too far from one another to be examined efficiently within one amplicon, such as with the attempt to simultaneously delete both -148kb to -137kb enhancer simultaneously, primer pairs were generated for individual guide cutting sites (as detailed in Table 2.4). Two strategies were used for the *in vitro* cutting assay with either one guide RNA per reaction, or a combination of a 5' and 3' guide RNA for a region. This enabled testing of whether guides could be used in conjunction with one another in order to excise a whole enhancer region *in vitro*. Where multiple guide RNA were used within a reaction, primer pairs were combined with the use of the forward primer from the 5' cutting substrate amplicon and reverse primer from the 3' cutting substrate amplicon.

Where guides were predicted to cut within enhancer regions, primers used for the general manipulation of the enhancer, such as those used for cloning (Chapter 2.3) were used.

Enhancer	Primer name	Sequence 5’ to 3’	T_m (°C)	Product length (bp)
-137kb 5’	137_5’C_F	GACCATATGAAGCGCAAGAAGG	59.8	529
	137_5’C_R	AATGGCCACTCCTCCTACTGA	60.3	
-137kb 3’	137_3’/2_C_F	AGCAAGGACATTCACTGGGCTC	62.1	978
	137_3’/2_C_R	GGAGGCGAAAACTGACTCCTGA	62.1	
-148kb 3’	148_3'/1_C_F	AAGCCTTGGGTTCCATCTGTGA	60.3	752
	148_3'/1_C_R	CCCAGAGAATGAGGAGCCAGAA	62.1	
-148kb/-137kb 5’	148-137_5'_C_F	CTACCCGTGTCTCCTTCTCCCA	64.0	884
	148-137_5'_C_R	GGACTGGTGAACCCTCATGGAC	64.0	
-148kb/-137kb 3’	148-137_3'_C_F	CAGTAGCCCAGACTAGCCCCAA	64.0	942
	148-137_3'_C_R	TCAGGTATCCAAAGAGCCTCGG	62.11	
-230kb 5’	230_5'_C_F	TTATTTATTTGCACGTTTTGCG	52.8	869
	230_5'_C_R	CCCACCAGAGGACAGTTAAAAG	60.3	
-230kb 3’	230_3'_C_F	CGCGCACATATCTATTGTTCAT	56.5	648
	230_3'_C_R	TAGCCCAGAGATTGGATGAAGT	58.4	

Table 2.4: Primers for the generation of substrate for in vitro CRISR cutting assay. Primers were designed to amplify products spanning CRISPR guide recognition sites

2.4.3.2 Substrate amplification and purification

Cutting assay substrate was generated using PCR. For amplicon shorter than 2kb in length GoTaq® G2 Flexi DNA Polymerase reaction protocol was used (Chapter 2.2.1.2). For amplicon greater than 2kb in length, Phire Hot Start II DNA Polymerase was used (Chapter 2.2.1.2). For all reactions 200ng of purified DNA template was used in a 50µL reaction.

PCR product specificity was assessed using agarose gel electrophoresis prior to purification using the Monarch® PCR and DNA Cleanup Kit (Chapter 2.2).

2.4.4 *In vitro* cutting assay

Incubation of Cas9 protein, guide RNA and DNA substrate enabled the capacity of the guide RNA to recruit Cas9 to target sites and therefore capacity to cut substrate to be assessed *in vitro*.

Streptococcus pyrogenes recombinant Cas9 protein containing a nuclear localisation signal (NLS) (Labomics/Toolgen, Cas9-TOO-50) was reconstituted in embryo water to concentration of 5µg/µL. For initial cutting assays, reaction conditions recommended by the Cas9 manufacturer were used and consisted of: 500ng Cas9 protein, 350ng of purified gRNA, 150ng of purified substrate, 1µL NEBuffer 3 (New England Biolabs), 1µL 10xBSA (New England Biolabs) and nuclease free water to a total volume of 10µL. 175ng of each guide RNA was used in duplex assays.

This reaction set up was altered to a molar ratio of 10:10:1 Cas9 protein: gRNA: substrate in optimised conditions (Gong *et al.* 2018a; Kouranova *et al.* 2016). Based on 500ng of Cas9 protein (molecular weight 160kDa) having molarity of 3.12µM, subsequent reactions were used with 3.12µM Cas9, 3.12µM gRNA and 312nM substrate DNA.

Molarity was determined on the basis that one base pair of nucleic acid has a molar mass of 660g/mol, with one nucleotide having a molar mass of 330g/mol

For example for 312nM of substrate with length of 1kb:

1000 base pairs of 660g/mol per base pair = 1000x660 = $6.6x10^5$

312nM = $3.12x10^{-13}$ mols/µL

$6.6x10^5$ x $3.12x10^{-13}$ = 2.06x10-7 g = 206ng substrate DNA

Reactions were incubated at 37°C for one hour prior to agarose gel electrophoresis. Control reactions were used for each substrate, with nuclease free water in place of guide RNA. This enabled the effect of gRNA on substrate to be more accurately assessed, and the extent to which presence of Cas9 protein slowed DNA fragment progression through agarose gel to be assessed.

2.5 Generation of transgenic mice

All animal work was conducted in accordance with the Animals (Scientific Procedures) Act (APSA) (1986) and was subject to local ethical committee review. This work was carried out under Professor George Bou-Gharios' project licence, PPL number 70/9047.

Mice were housed at the University of Liverpool's Biomedical Services Unit under specific pathogen free (SPF) environmental conditions. The mice were housed with maintained at 22±2 ˚C and a cycle of 12 hours of light, 12 hours of darkness.

All purchased wild-type mice were procured from Charles River (Wilmington, Massachusetts, United States). Transgenic mice lines were generated using B6CBAF1 as a background strain. These mice are generated as F1 cross of pairing female C57BL/6J and male CBA/CaCrl.

Aside from preparation of DNA constructs and guide RNA, each stage of this protocol was carried out by Dr Ke Liu (University of Liverpool). Microinjection of cell cell-stage embryos was carried out in accordance with Ittner and Götz (2007).

2.5.1 Reporter construct DNA preparation

Plasmids containing enhancer regions driving reporter gene via a promoter region (as detailed in Chapters 2.3 and 5.2) were cut in order to produce constructs for examination of *in vivo* reporter gene expression, as schematically represented in

Figure 2.11. Each transgenic line was named in accordance with the plasmid the construct originated from.

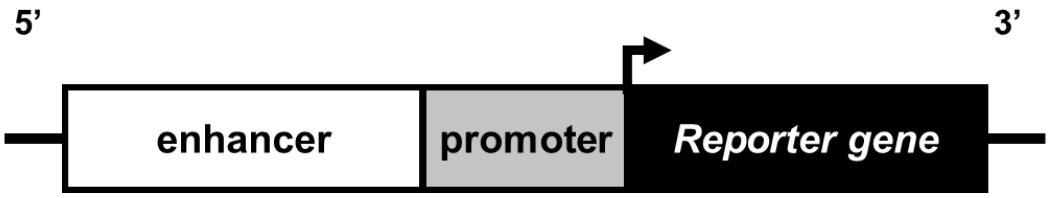


Figure 2.11: Schematic representation of DNA constructs used in the creation of transgenic mice. Constructs consisted of enhancer regions (white) placed upstream of a promoter (grey) and reporter gene (black), with the enhancer therefore regulating the transcription of the reporter gene.

LacZ containing plasmids were cut using restriction enzymes recognising sequences upstream of the 5' enhancer sequence boundary and downstream of the 3' polyA tail of the *lacZ* gene. This ensured the removal of sequences necessary for bacterial maintenance that are not required in mammalian *in vivo* system. Plasmid DNA was prepared as outlined in Chapter 2.3. Restriction digests were carried with agarose gel electrophoresis and cutting of the gel in order to isolate DNA fragments of interest (outlined in Chapter 2.2 and illustrated in Chapter 5.2).

In order to create the -148kb_short Hsp68LacZ transgenic mice, the -148kbCTGFHsp68LacZGW plasmid was cut using ScaI, which cuts 1044bp downstream of the 5' enhancer boundary, leading to a 1531bp sub-region of the enhancer driving reporter gene expression.

Construct DNA fragments were purified using the gel extraction protocol as outlined in Chapter 2.2.3. However, after the initial centrifugation step following addition of sample to the column, 500µL buffer QG was added to the column before centrifugation for one minute and continuation of standard protocol. Purified DNA was pooled through elution from multiple QIAquick columns to a total of 30µL in embryo water. Constructs were further purified by being placed on an Ultrafree®-MC column (Merck Millipore, UFC30GV25) which was subsequently centrifuged for two minutes at 12,000*g*.

Concentration of construct DNA was determined using gel electrophoresis with 0.8% agarose gel containing 0.5µg/µL Ethidium Bromide, with sample dilutions ran alongside 0.5µg and 1.0µg each of λ-HindIII digest and 2 log ladders. Concentration of construct DNA was estimated based on the known mass of ladder bands.

2.5.2 Preparation of embryos

Female B6CBAF1 mice aged five to six weeks were treated to induce superovulation through intraperitoneal injection (IP) of 5U follicle stimulating hormone (FSH) on day one, followed by IP injection of 5U luteinising hormone (LH) on day three. These mice were paired with stud male B6CBAF1 overnight. The following morning embryos were isolated (0.5 days post-coitum (dpc)) from females where copulation plugs were present. 300µg/mL hyaluronidase was then used to remove cumulus cells before zygotes were transferred to M16 media (Merck Millipore, MR-016-D). This procedure was carried out by Dr Ke Liu.

2.5.3 Microinjection

For reporter gene constructs, DNA was diluted to 2ng/µL in embryo water before injection into the pro-nuclei of viable embryos.

For the generation of mice harbouring CRISPR-Cas9 based genomic alterations, guide RNA was incubated with Cas9 protein at an equimolar ratio (either 0.5µM or 1.5µM) for 10 minutes at room temperature prior to the commencement of injection to allow protein-gRNA complex formation (Jinek *et al.* 2012). Recombinant Cas9 protein was prepared as in Chapter 2.4.3. Complexes were injected into the cytoplasm of embryos as the Cas9 protein used has a nuclear localisation signal (NLS) (Horii *et al.* 2014). Injected embryos were incubated in M16 media until transfer. This procedure was carried out by Dr Ke Liu.

2.5.4 Embryo transfer

Female CD1 mice were paired with vasectomised male CD1 mice overnight. Where copulation plug had formed, resultant pseudo-pregnant females were used as surrogate mothers for embryo transfer. Female mice were anaesthetised before 10-15 viable embryos at the one cell stage that had survived microinjection were transferred to the ampulla of uterine tube, approximately 30 embryos per female mouse. This procedure was carried out by Dr Ke Liu.

2.5.5 Genotyping

2.5.5.1 DNA extraction

Ear notch samples were taken from mice upon weaning at approximately three weeks of age. Samples for genotyping were also taken from the distal portion of tails, or ears of mice collected for tissue, in addition to placental tissue from embryos.

Samples were suspended in 100µL lysis buffer (50mM Tris-HCl, pH 8.0; 0.1M NaCl; 1% w/v SDS; 20mM EDTA) before the addition of 10µL 10mg/mL Proteinase K (Sigma Aldrich, P2308 or Bioline, BIO-37037) (stored in 50% glycerol solution). Samples were then vortexed briefly before incubation overnight at 55°C.

Following incubation overnight, samples were allowed to cool to room temperature before being briefly vortexed. Samples were examined to ensure that tissue had been broken down in the digestion step. Samples were then centrifuged at 16200*g* for three minutes. Resultant supernatant was transferred to clean 1.5mL microcentrifuge tube before the addition of an equal amount of molecular grade absolute isopropanol. Samples were gently inverted twice before centrifugation using aforementioned conditions. Supernatant was discarded before samples were centrifuged at 16200*g* for a further 30 seconds, after which supernatant was removed with a pipette. Pellets were left to air-dry at room temperature for approximately ten minutes. 25µL TE buffer was subsequently used to reconstitute pellets, samples were gently flicked in order to aid this before incubation at room temperature for one hour with gentle flicking of sample every 15 minutes in order to aid reconstitution.

This protocol was later adapted with an ethanol pellet wash step in order to reduce isopropanol contamination and therefore improve ratio for 260/230nm absorbance measurement for DNA samples. Prior to the pellet drying step, 500µL 70% (v/v) molecular grade ethanol in nuclease free water was added to each sample before being discarded and pellet left to dry prior to standard reconstitution method.

For adult tail samples and embryonic placental material, double volumes of each reagent were used at each stage.

Concentration and quality of resultant DNA was assessed using Nanodrop as aforementioned.

2.5.5.2 Genotyping PCR

An internal control PCR reaction was used for each DNA sample to ensure DNA purity and integrity was sufficient for PCR and to allow determination of false-negatives in amplicon of interest PCR reactions. The internal forward primer sequence was 5'- TGG ACA GGA CTG GAC CTC TGC TTT CCT AGA and the internal control reverse primer had sequence of 5'- TAG AGC TTT GCC ACA TCA CAG GTC ATT CAG. The internal control reaction amplicon is a 194bp product which spans exon 1 of the Fabp2 gene (chr3:122,598,282-122,598,475). For the CRISPR genome edited transgenic mice, this locus was not predicted to be affected by off-target effects of any of the guide RNAs. Therefore, this reaction would still be suitable as a control in the genotyping of CRISPR transgenic mice.

GoTaq® G2 Flexi DNA polymerase (as outlined in Chapter 2.2.1.2) was used to amplify products of interest with 100ng of template DNA in each reaction with primers listed in Table 2.5.

Construct specific primers were designed which spanned from the enhancer region into the promoter region. Preliminary PCR using purified genomic DNA as a template was used to check primer specificity and optimise amplification conditions for each primer pair used prior to usage with DNA samples extracted for genotyping.

Where primer T_m similarity allowed, duplex reactions were used whereby the internal control and amplicon for genomic alteration of interest were amplified within one reaction. PCR reaction set-up was modified for this approach as follows: 5µL 5x reaction buffer, 2µL 25mM $MgCl_2$, 0.75µL 10mM dNTP, 0.5µL 10µM forward internal reaction primer, 0.5µL 10µM reverse internal reaction primer, 0.5µL 10µM forward amplicon of interest primer, 0.5µL 10µM reverse amplicon of interest primer, 100ng template, 0.15µL GoTaq® G2 Flexi DNA polymerase with water to a total of 25µL per reaction.

Where copy number was trying to be assessed or DNA impurity hampered the efficiency of reactions, template amount was reduced to 50ng or 25ng per reaction.

Amplicon	Primer name	Sequence 5' to 3'	T_m (°C)	Product length (bp)
LacZ reporter gene	LacZ 008	GTTGCAGTGCACGGCAGATACACTTGCTGA	69.5	389
	LacZ 009	GCCACTGGTGTGGGCCATAATTCAATTCGC	69.5	
CTGF-4 (-4kbCTGFHsp68LacZ)	-4CTGFGWgeno_F	CAAAAAGGTTGGGCTAGGGC	59.4	550
	148CTGFgeno2_R	GTCCGGTGACGTGATCCTCT	61.3	
CTGF137s (-137kbshortCTGFHsp68LacZ	-137sGWgeno_F	GACTGCTGAAGACGTCGCTACAT	62.4	608
	148CTGFgeno2_R	GTCCGGTGACGTGATCCTCT	61.3	
CTGF148Hsp68lacZ (CTGF148 and -148kb_short)	148CTGFgeno2_F	GCACTCTTCCCAATATTTATGTATGC	58.7	749
	148CTGFgeno2_R	GTCCGGTGACGTGATCCTCT	61.3	
CTGF230	230GWgeno_ F	GAGTCGCATAAATACCTTGTGAGC	61.0	789
	148CTGFgeno2_R	GTCCGGTGACGTGATCCTCT	61.3	

Table 2.5: Primers used to genotype *LacZ* reporter transgenic mice. For construct specific genotyping, PCR amplicon were selected spanning from the enhancer region into the promoter region. Further primers were also used that amplified within the *LacZ* gene

The 148CTGFgeno2_R primer was used as reverse primer for all construct specific reactions for *lacZ* reporter lines as this primer sits within the Hsp68 promoter region. For the CRIPSR-Cas9 transgenic mice lines, primers used for generating substrate for the in vitro cutting assay were used to genotype using the reaction set up and conditions outlined in appendices Chapter 5.3. Amplification of these regions would not be possible if the enhancer regions were excised.

PCR products were ran on 1% agarose gels before visualisation for confirmation of genotype.

2.6 X-gal (5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside) staining

2.6.1 X-gal staining

Expression of *lacZ* reporter gene expression was qualitatively assessed using X-gal staining.

All mice taken for tissue were culled in accordance with Home Office protocol and housed as aforementioned in Chapter 2.5.

For embryonic time points, mice were paired for two evenings with presence of copulation plug in the morning assumed to be E0.5. Embryos were taken at E11.5, E12.5, E13.5, E15.5 and E17.5 dpc. Embryos were placed in cold phosphate buffered saline (PBS) before removal of placenta and yolk sac. For E17.5 embryos, skin was dissected in addition to a sagittal abdominal incision to allow penetration of solutions. For the -4kb, -102kb, -137kb and preliminary work for -148kb transgenic assay at E15.5, Christoph Zimmer or Dr Ian Li carried out this procedure.

Adult tissue samples were also placed in cold PBS once dissected. Skeletal samples taken consisted of the cranium, thoracic vertebrae, lumbar vertebrae, ribs and hind limbs. Soft tissue samples were taken from the brain, eye, heart, lung, kidney and liver. For knee tissue samples, incisions were made within the menisci region and dorsal aspect of knee in order to destabilise the joint and allow solution penetration into the medial regions of the joint.

All samples were rinsed through placing them in fresh cold PBS, prior to transfer to cold X-gal fix solution (0.2% glutaraldehyde solution (v/v), 0.1 M sodium phosphate

buffer pH 7.3, 5mM Ethylene glycol-bis(2-aminoethylether)-N,N,N′,N′-tetraacetic acid (EGTA) pH 8.0, 2mM $MgCl_2$, 2% neutral buffered formalin) on ice. Samples were then fixed at room temperature whilst rotating for a duration of time dependent on developmental stage (Table 2.6)

Developmental stage	Fixation duration
E11.5	20 minutes
E13.5	35 minutes
E15.5	45 minutes
E17.5	60 minutes
adult	120 minutes

Table 2.6: X-gal staining tissue fixation durations. Fixation duration depended on developmental time-point

After fixation, fix solution was decanted and replaced with rinse solution (0.1M sodium phosphate buffer, 2mM $MgCl_2$, 0.1% sodium deoxycholate (w/v), 0.2% NP-40 substitute (v/v). Samples were rinsed initially whilst rotating at room temperature for 30 minutes before rinse solution was discarded. This process was repeated a further two times. X-gal stain solution, composed of: 1mg/mL X-gal substrate (VWR, 437132J), 5mM potassium ferricyanide (Sigma Aldrich, P-3667), 4mM potassium ferrocyanide (Sigma Aldrich, P-9387) dissolved in rinse solution was then added to samples which were then shielded from light before rotating incubation overnight at room temperature. Progression of staining was observed at regular intervals, typically hourly for the first four hours. After staining period, stain solution was discarded before addition of rinse solution and three stages of rinsing as carried out previously.

Whole mount imaging was conducted using an Olympus SZX12 microscope with QCapture camera and software (Q Imaging, Surrey, British Columbia, Canada).

2.6.2 Soft tissue clearing of X-gal stained tissue

The soft tissue was cleared from X-gal stained embryos and adult skeletons in preparation for further whole mount imaging, based on the protocol of Rigueur and Lyons (2014). This was carried out after post staining-fix in 10% neutral buffered formalin solution for 24 hours at 4°C. Embryos were firstly transferred to a 1% (w/v)

potassium hydroxide (KOH) solution before being left at room temperature overnight. This solution was replaced daily with repetition of incubation process for three days. After this point, the solution was changed to 0.8%KOH/20% glycerol (Sigma Aldrich, G5516), after which samples were left at room temperature for seven days. This incubation process was repeated firstly with 0.6%KOH/40% glycerol and subsequently 0.2%KOH/80% glycerol. Whole mount images were taken before embryos were stored in 100% glycerol.

For adult skeletal samples, as much soft tissue such as muscle was removed through dissection, with ligamentous tissue left intact in order to maintain articulation of the skeleton. Samples were transferred to 2% KOH solution for a total of seven days at room temperature, with total solution change daily. After this period, solution was changed to 1% KOH with replacement every 48 hours or when solution became discoloured. As with the embryos, the ratio of KOH in clearing solution was gradually reduced, with increasing proportion of glycerol. Samples were imaged and stored in 100% glycerol.

The background was removed from images using Gnu Image Manipulation freeware software.

2.7 Histology

Tissue samples were fixed at 4°C for a period of 24 hours using 10% neutral buffered formalin solution. Samples were then placed in 70% ethanol solution until tissue processing. Skeletal tissue samples were decalcified through being placed in 10% EDTA (w/v) pH 7.4 solution on a rocking platform at room temperature for 2 weeks prior to processing.

2.7.1 Soft tissue clearing

All tissue samples were processed using Leica ASP300 tissue processor.

For soft tissue samples and embryos, the following parameters were used:

Solution	Duration	Temperature
Ethanol (70%)	30 minutes	Room temperature
Ethanol (90%)	30 minutes	Room temperature
Ethanol (absolute)	5 minutes	Room temperature
	10 minutes	Room temperature
	10 minutes	Room temperature
	20 minutes	Room temperature
Xylene	10 minutes	Room temperature
	20 minutes	Room temperature
	30 minutes	40°C
Wax	30 minutes	62°C
	60 minutes	62°C
	150 minutes	62°C

For hard tissue; decalcified skeletal samples, the following parameters were used:

Solution	Duration	Temperature
Ethanol (70%)	15 minutes	Room temperature
Ethanol (90%)	60 minutes	Room temperature
Ethanol (absolute)	20 minutes	Room temperature
	40 minutes	Room temperature
	60 minutes	Room temperature
	120 minutes	Room temperature
Xylene	20 minutes	Room temperature
	40 minutes	Room temperature
	60 minutes	40°C
Wax	30 minutes	62°C
	60 minutes	62°C
	150 minutes	62°C

After processing, samples were embedded within Surgiplast Paraplast (Leica, 39601006) using Leica EG1150 C embedding station.

Sections of samples were cut to a thickness of six micrometres (μm) using a Leica 2135 manual microtome. All embryos were sectioned in the sagittal plane.

Postnatal tissue samples were sectioned in the coronal plane aside from the paw which was sectioned in the transverse plane. Sections were transferred to slides which were incubated at 37°C overnight.

2.7.2 Eosin counterstaining and imaging

Samples stained for β-galactosidase activity were counterstained using eosin. Briefly, slides were dewaxed through two stages of being placed in Histo-Clear (National Diagnostics, HS-200) for two minutes. Slides were then transferred to 100% ethanol for two minutes, then 90% ethanol solution for two minutes, then 70% ethanol solution for two minutes. Slides were then dipped in running tap water before being transferred to eosin Y (alcoholic eosin, Leica, 3801600BBE) for two minutes. Slides were then dipped in running tap water before dehydration through being placed in 70% ethanol for 30 seconds, 90% ethanol for 30 seconds and 100% ethanol for 30 seconds. Slides were then transferred to Histo-Clear. Specimens were mounted and coverlipped using Pertex® (Histolab, 00811) or DPX (BDH, 360294H) as mounting media.

Specimens were imaged using an Olympus BX60 light microscope in conjunction with a Zeiss Axiocam camera system using ZEN blue edition software. The Atlas of Mouse Development (Kaufman 2003) and eHistology Atlas (http://www.emouseatlas.org/emap/eHistology) (Graham *et al.* 2015) were used to identify histological tissues.

2.8 In vitro modelling of enhancer function

2.8.1 pRL-TK

Immortalised cell lines were used to model enhancer function. Cells were transfected with firefly luciferase reporter plasmid DNA; the creation and preparation of which is outlined in Chapter 2.3.

The pRL-TK plasmid (Promega, E2241) was used as an internal control in luciferase assays. This vector encodes a *Renilla* luciferase gene, the expression of which is driven by a thymidine kinase (TK) promoter region taken from Herpes Simplex Virus for consistent basal expression of *Renilla* luciferase, as illustrated in Figure 2.12.

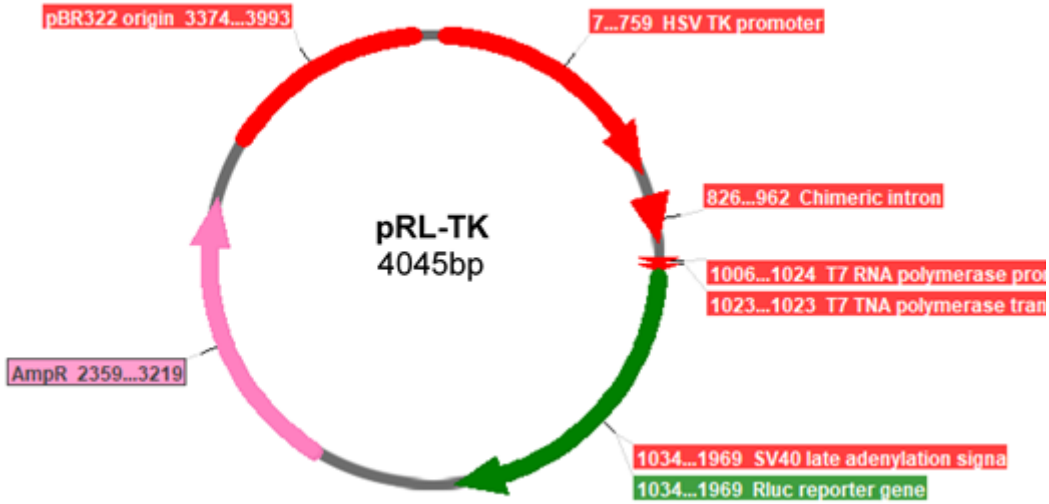


Figure 2.12: Map of pRL-TK expression plasmid. This expression plasmid contains a renilla luciferase gene (*Rluc*), the expression of which driven by a thymidine kinase (*TK*) promoter to give low to moderate levels of gene expression.

Co-transfection of Firefly luciferase and *Renilla* luciferase vectors in conjunction with use of the Dual-Luciferase® Reporter Assay System (Promega, E1910) enables signal stemming from Firefly luciferase driven by enhancer regions to be distinguished from *Renilla* luciferase from pRL-TK. This therefore enables discrepancies in cell number or transfection efficiency to be overcome in assessing enhancer function.

2.8.2 Cell culture

All cells were maintained in incubators with 5% CO_2 at 37ºC. Cells were regularly passaged when confluence reached 70%-90% at constant split ratio; typically 1:3. All cells were grown in media supplemented with 10% heat inactivated Foetal Bovine Serum (FBS) (ThermoFisher Scientific, 10500064) unless otherwise stated. During passage and seeding, cells were detached from culture plastic using TypLE™ Express Enzyme (ThermoFisher Scientific, 12604).

HTB-94 human chondrosarcoma cells (also known as SW-1353), Human Embryonic Kidney cells (HEK293-A) and NIH3T3 murine embryonic fibroblast cells were maintained in high glucose Dulbecco's Modified Eagle Medium (ThermoFisher Scientific, 41965039). TC28-i2 human rib derived chondrocyte cells were maintained using DMEM/Ham's F-12 (ThermoFisher Scientific, 11320). Murine calvarial MC3T3-E1 were cultured using α- Modified Eagle Medium (MEM) (no nucleosides, no ascorbic acid; ThermoFisher, A1049001).

2.8.3 Transfection of cells

Transfections were carried out in 48 well plate format. Cells were seeded 24 hours prior to transfection, with a density of $1.75x10^4$ cells per well for each cell line.

Triplicate reactions were carried out for each reporter plasmid of interest in each transfection. Lipofectamine® 2000 (ThermoFisher Scientific, 11668019) was used as transfection reagent. For each well transfected, 500ng of firefly reporter plasmid, 30ng of renilla plasmid and 1µL Lipofectamine 2000® was used. Mastermixes were created for each pSLF01 vector variant of interest. Firstly, firefly and renilla plasmid DNA was diluted in Opti-MEM™ with gentle aspiration to mix (ThermoFisher Scientific, 31985070). Lipofectamine® was also diluted in Opti-MEM™ before resultant solution was gently flicked to mix and incubated at room temperature for five minutes. Opti-MEM dilution solutions were then combined in 1:1 ratio before being gently flicked to mix and incubation at room temperature for 20 minutes. Resultant transfection complexes were added dropwise to the centre of seeded cell wells, before plates were returned to 37°C incubator. After four hours of incubation standard media was added to each transfected well.

Control procedure were also carried out whereby lipofectamine and Opti-MEM™ mixture was added to the cells in the same manner as DNA transfection complexes.

2.8.4 Luciferase assay

2.8.4.1 Cell lysis

For each stage of the luciferase assay, reagents from the Dual-Luciferase® kit were used in accordance with the manufacturer's instructions.

24 hours after transfection, the media was removed from cell culture plate wells and cells were washed with PBS. This wash solution was removed before cells were lysed using Passive Lysis Buffer that had been diluted to 1x working solution using distilled water. 65µL of Passive Lysis Buffer was added to each well, after which plates were left a rocking platform for 15 minutes at room temperature. Resultant lysed cell solution was transferred to a clean microcentrifuge tube.

2.8.4.2 Luminescence reading

Luminescence was assessed using Glomax® Multi-Detection System (Promega). 50µL of Luciferase Assay Reagent II (LARII) was added to triplicate wells within a white walled plate (Sigma Aldrich, M4811) for each pSLF01 variant sample. Luminescence in the plate was then read with two second integration period per well using the luminescence program and recorded as a Microsoft Excel file. This therefore allowed the identification of any background luciferase activity in the plate or reagent. 10µL of cell lysate was then added to each well, with the contents being gently aspirated four times in order to ensure samples were thoroughly mixed but no bubbles were present. Luminescence was then measured again, in the same manner as before with generation of another spreadsheet containing raw data. 50µL Stop and Glo® reagent was then added to each well, before gentle aspiration and reading of luminescence and collection of excel data file as before.

2.8.4.3 Data interpretation

The luminescent signal produced by the firefly luciferase protein and therefore pSLF01 plasmid variant function was the signal detected after the addition of LAR II. The data point from each well was divided by that of the reading from the same well after the addition of Stop and Glo® (renilla luciferase signal). This allowed normalisation of luciferase signal and accounted for any discrepancies in cell number between wells. Means were taken for the normalised data for each pSLF01 variant.

Fold differences in luciferase expression were calculated and normalised to the basic promoter along pSLF01 plasmid. Data was assumed to be normally distributed and unpaired Student's *t*-tests were carried out using Microsoft Excel to compare responses between enhancer regions.

2.9 Identification and in vitro validation of transcription factor binding sites

2.9.1 Putative transcription factor binding motif prediction

2.9.1.1 Motif prediction

The Transcription Factor Affinity Prediction (TRAP) web tool, available at: http://trap.molgen.mpg.de/cgi-bin/home.cgi was used to identify putative TFBS. The results generated from this tool are based on affinity between transcription factor and DNA sequence. TF are ranked based on the probability scores (*p*-value) for affinity to the sequence compared with distribution within the background model. *p*-values of <0.05 correspond to predicted transcription factor-sequence of interest affinity which exceeds that of the transcription factor-random sequence within the background model. Therefore a *p* value <0.05 dictates significant TF-sequence of interest interaction (Manke *et al.* 2008).

The TRAP single sequence function was used (Manke *et al.* 2008; Roider *et al.* 2007). Sequences for regions of interest were entered in FASTA format, with testing against the TRANSFAC database of TFBS ('transfac_2010.1' vertebrate matrix file), mouse promoters as the background model and Benjamini-Hochberg multiple test correction (Figure 2.13).

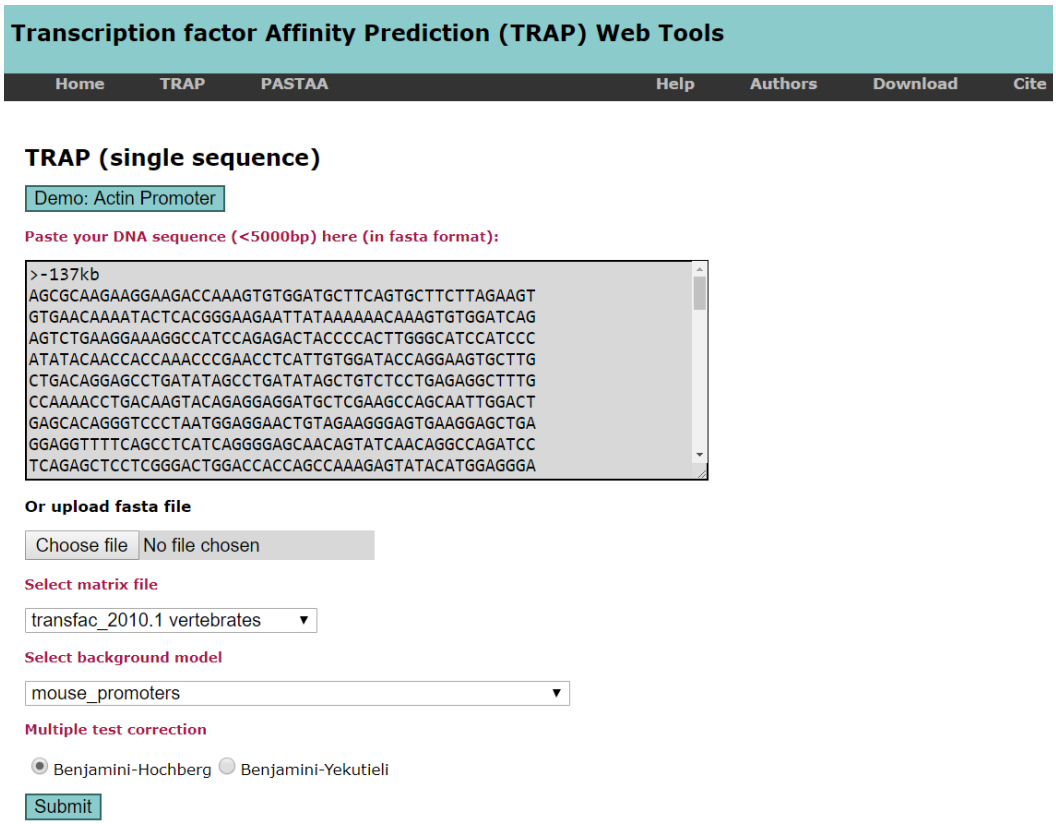
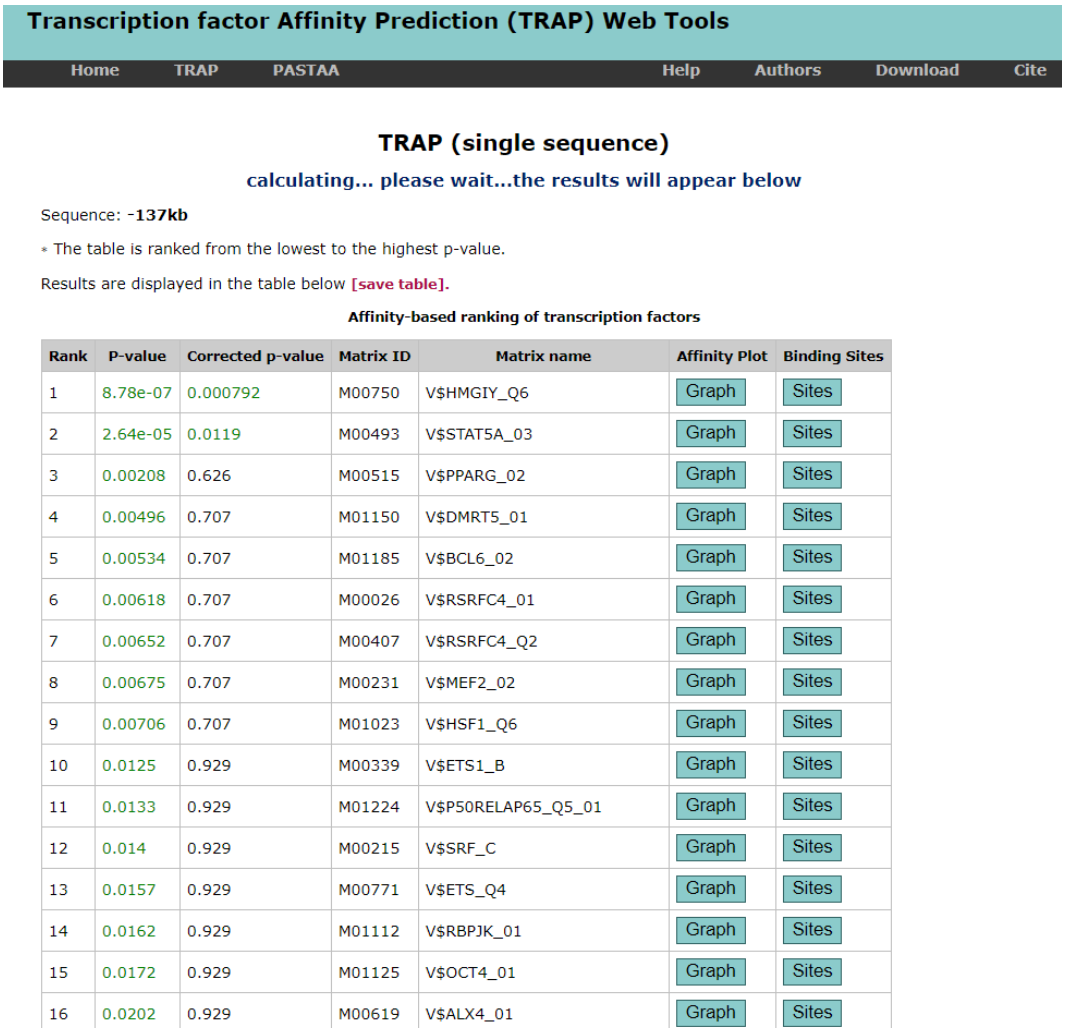


Figure 2.13: Screenshot of entry of DNA sequences into the TRAP tool. Sequences were entered in FASTA format, with results generated using the transfac_2010.1 vertebrate matrix file and mouse promoter as background model. Benjamini-Hochberg was selected for multiple test correction

TRAP results are obtained as a ranked list of transcription factors, with *p*-value, matrix number and matrix name (Figure 2.14 and Chapter 5.4). The matrix number is a unique to each transcription factor. There may be multiple matrices per TF, therefore each matrix number reflects a specific consensus binding motif. The positioning of binding sites can also be visualised in the results using the Binding Sites 'sites' function.



Transcription factor Affinity Prediction (TRAP) Web Tools

Home TRAP PASTAA Help Authors Download Cite

TRAP (single sequence)

calculating... please wait...the results will appear below

Sequence: **-137kb**

- The table is ranked from the lowest to the highest p-value.

Results are displayed in the table below [save table].

Affinity-based ranking of transcription factors

Rank	P-value	Corrected p-value	Matrix ID	Matrix name	Affinity Plot	Binding Sites
1	8.78e-07	0.000792	M00750	V$HMGIY_Q6	Graph	Sites
2	2.64e-05	0.0119	M00493	V$STAT5A_03	Graph	Sites
3	0.00208	0.626	M00515	V$PPARG_02	Graph	Sites
4	0.00496	0.707	M01150	V$DMRT5_01	Graph	Sites
5	0.00534	0.707	M01185	V$BCL6_02	Graph	Sites
6	0.00618	0.707	M00026	V$RSRFC4_01	Graph	Sites
7	0.00652	0.707	M00407	V$RSRFC4_Q2	Graph	Sites
8	0.00675	0.707	M00231	V$MEF2_02	Graph	Sites
9	0.00706	0.707	M01023	V$HSF1_Q6	Graph	Sites
10	0.0125	0.929	M00339	V$ETS1_B	Graph	Sites
11	0.0133	0.929	M01224	V$P50RELAP65_Q5_01	Graph	Sites
12	0.014	0.929	M00215	V$SRF_C	Graph	Sites
13	0.0157	0.929	M00771	V$ETS_Q4	Graph	Sites
14	0.0162	0.929	M01112	V$RBPJK_01	Graph	Sites
15	0.0172	0.929	M01125	V$OCT4_01	Graph	Sites
16	0.0202	0.929	M00619	V$ALX4_01	Graph	Sites

Figure 2.14: Selection of transcription factor binding matrices generated by TRAP. Transcription factor matrices are ranked based on the probability of interaction (P-value). Results can be interpreted through the affinity plot and binding site functions which allow visualisation of TFBS within sequence of interest.

A specific matrix ID is selected and the DNA sequence of the region is visualised with highlighting of the sequences predicted to be bound by the transcription factor. The distribution of sites for a matrix across a region of interest can also be visualised using the Affinity Plot 'graph' function.

MotifMap (http://motifmap.ics.uci.edu) (Daily *et al.* 2011) and SwissRegulon Portal (http://swissregulon.unibas.ch/sr/) (Pachkov *et al.* 2013) are both online databases of transcription factor binding motifs. These resources were used as to find documented consensus motifs related to the TFBS within sequences of interest.

2.9.1.2 Examination of TFBS sequence conservation

The conservation of putative TFBSs were assessed using the aforementioned Multiz function within the UCSC Genome Browser (Chapter 2.1), in addition to Clustal Omega. Firstly sequences of interest were identified within the UCSC Genome Browser, 'show details for feature' was then used to examine aligned sequences from organisms including; mouse, rat, human, opossum, platypus, chicken, lizard and fugu (Figure 2.15)

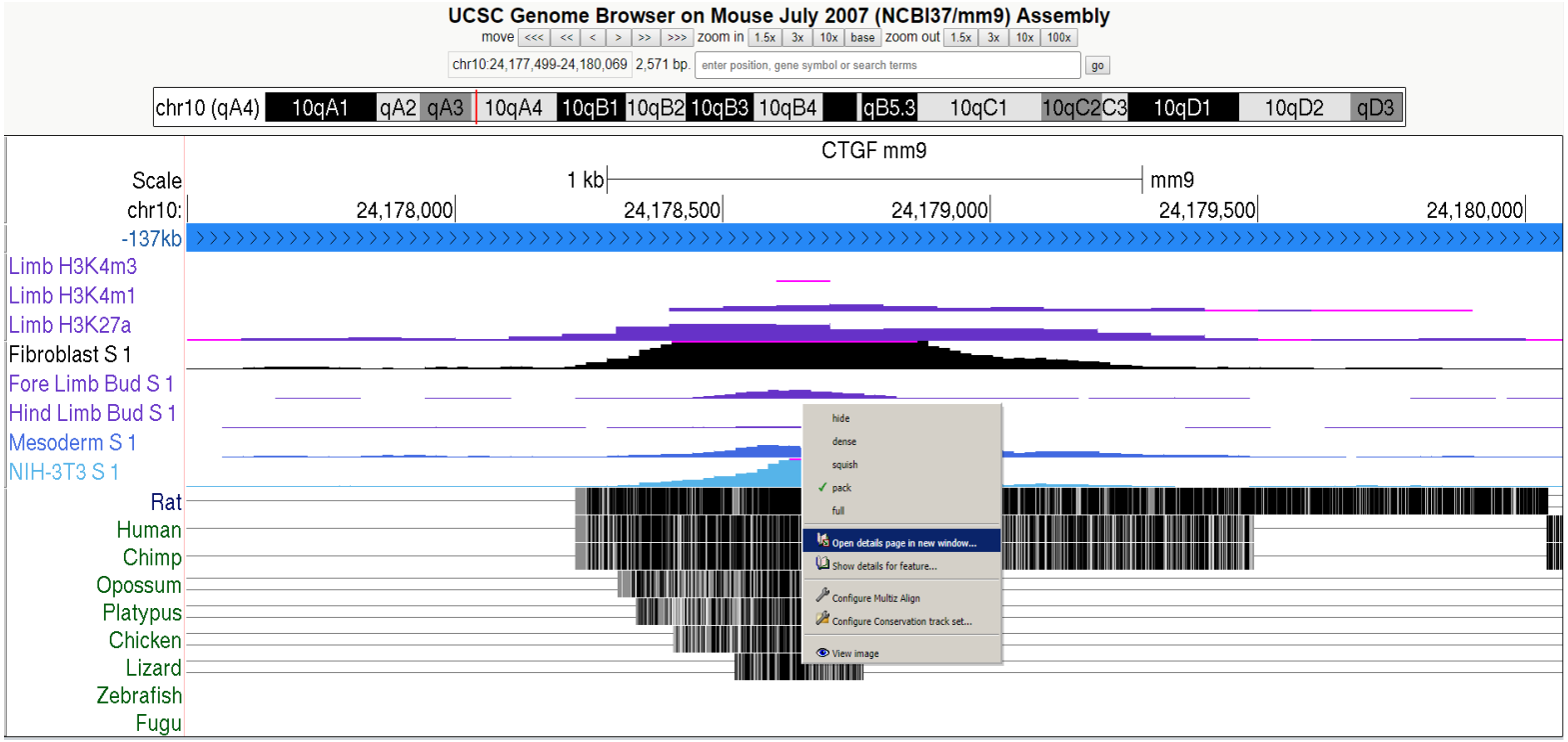


Figure 2.15: UCSC ENCODE Genome Browser based identification of conserved sequences. Alignment of sequences from multiple organisms was visualised with the Multiz alignment tracks. Sequences similar between species are represented by black bars. By clicking on conserved sequences within the Genome Browser, alignment at nucleotide level can be accessed.

This enabled the base-wise conservation of sites of interest between organisms to be visualised, as demonstrated in Figure 2.16.

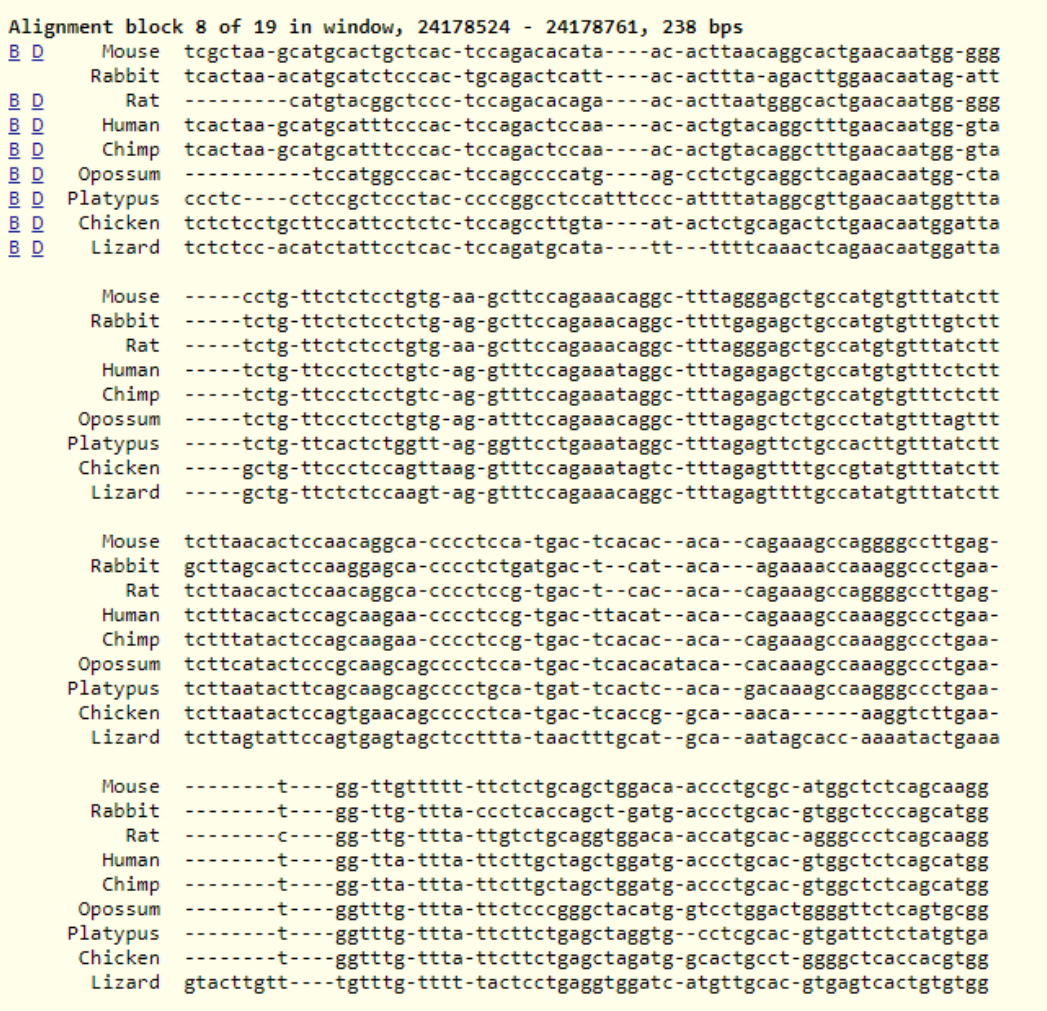


Figure 2.16: Visualisation of Multiz alignment of conservation of DNA sequences. Nucleotides are organised into blocks which are aligned based on sequence conservation across species of interest.

DNA sequences were then taken from Multiz for mouse, human and the most evolutionarily distant, yet conserved organism and aligned using Clustal Omega (Sievers *et al.* 2011), available at https://www.ebi.ac.uk/Tools/msa/clustalo/. Sequences were inputted in FASTA format, with output format of ClustalW with character counts as shown in Figure 2.17. Reverse complement function was used to ensure that DNA for organisms; for example humans, where sequence running along negative strand was efficiently aligned against that of the mouse.

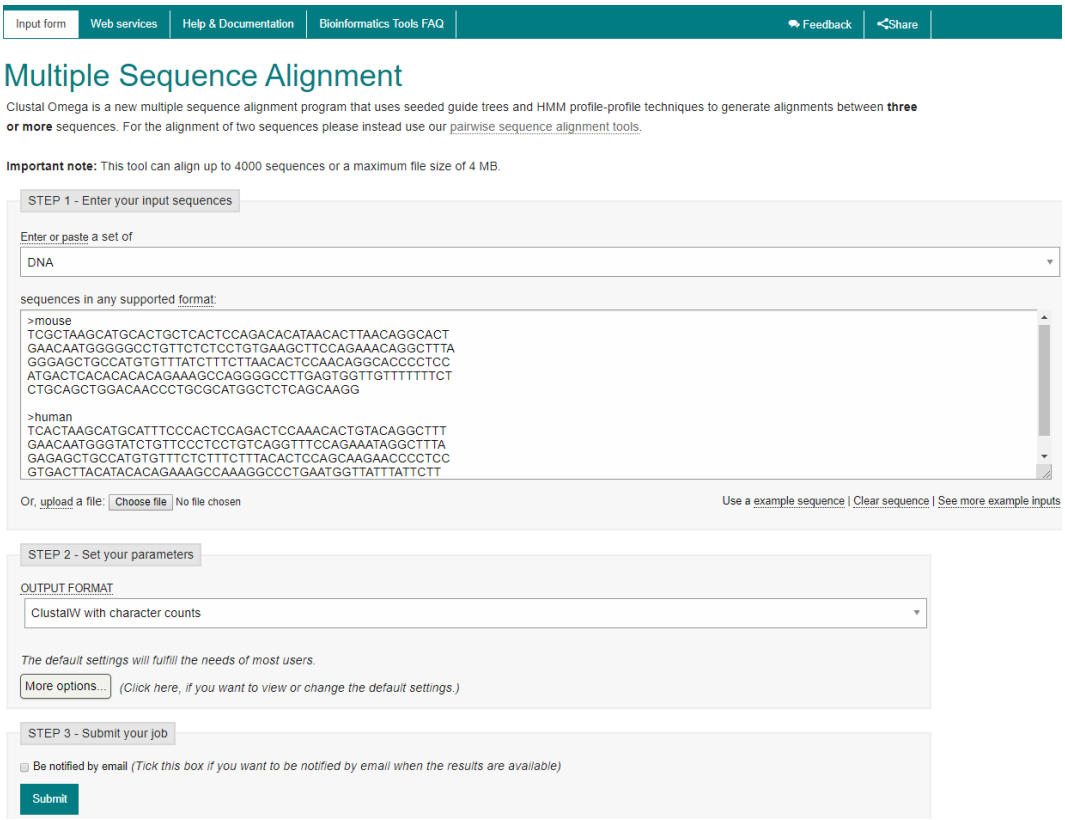


Figure 2.17: Input of DNA sequence from various organisms into the ClustalOmega interface. Sequences from each organism of interest gathered from Multiz (Figure 2.16) were entered in FASTA format and aligned using the ClustalW function.

Results are a percentage of conserved bases between the aligned sequences, in addition conservation of aligned between sequences can be visualised with asterisk denoting aligned nucleotides between all species under investigation.

2.9.2 Electrophoretic mobility shift assay (EMSA)

EMSA methodology employed herein was adapted from that of Dr Louise Reynard at Newcastle University.

2.9.2.1Oligonucleotide probe and competitor design

Oligonucleotides were designed to span putative TFBS identified with additional flanking sequence to a total of approximately 30 nucleotides in length. Oligonucleotide sequences were inputted into TRAP in order to ascertain specificity of probe to transcription factor of interest and verify predicted consensus motif positioning within probe. For mutant oligonucleotides, nucleotides within the consensus binding motif were substituted in order to limit recognition and interaction with the cognate transcription factor. Mutant oligonucleotide sequence

TF affinity was compared to that of the WT oligonucleotide using the single sequence TRAP interface as aforementioned.

2.9.2.2 Oligonucleotide preparation

Forward and reverse complement oligonucleotides were custom synthesised and purchased for each TFBS of interest, incorporating a 5' dye at each end with high performance liquid chromatography (HPLC) grade purification. Oligonucleotides purchased from Integrated DNA Technologies had 'IRDye™700' as the 5' modification, whereas oligonucleotides procured from Eurofins Genomics had a 5' 'DY682' modification. Both modifications were compatible with the Li-Cor Odyssey® imaging system (Li-Cor, Lincoln, Nebraska, United States) Light exposure was minimised at all times in order to prevent bleaching of probes.

For RUNX2 optimisation a positive control probe was used based on validated RUNX2 binding site within the *osteocalcin* gene promoter sequence (Lamour *et al*. 2007). The probe contained DY682 modification at the 5' end. Forward probe sequence: 5'- GCTGCAGTCACCAACCACAGCATCCTTTGG and reverse probe sequence 5'- CCAAAGGATGCTGTGGTTGGTGACTGCAGC.

Sequences for probes and competitors are detailed in Tables 2.7 to Table 2.9.

Lyophilised probes were reconstituted to 100μM master stock solution using nuclease free water and stored at -80°C. Probes were annealed as follows; 10μL each of forward and reverse 100μM stock were added to 10μL annealing buffer (100mM Tris HCl, 500mM NaCl, 10mM EDTA) along with 70μL nuclease free water, before incubation at 95°C for five minutes in a heat block, which was allowed to cool to room temperature before removal of resultant 20μM annealed probe stocks. Probes were further diluted to 100nM stock using nuclease free water, and stored at -20°C.

Competitor oligonucleotides were purchased for each probe of interest (Eurofins Genomics) and reconstituted to 100μM with nuclease free water. These oligonucleotides were annealed in the same manner as for the probes.

Probe/ competitor	Sequence name	label	Sequence 5’ to 3’
Probe	137sox9WT_F	IRD700	GGCACTGAACAATGGGGGCCTGTTCTCTCCTGT
Probe	137sox9WT_R	IRD700	ACAGGAGAGAACAGGCCCCCATTGTTCAGTGCC
Probe	137sox9MUT_F	IRD700	GGCACTGCCCGGTGGGGGCCTGTTCTCTCCTGT
Probe	137sox9MUT_R	IRD700	ACAGGAGAGAACAGGCCCCCACCGGGCAGTGCC
Competitor	137SOX9CTRL_F	none	CTAGGGCACTGAACAATGGGGGCCTGTTCTCTCCTGT
Competitor	137SOX9CTRL_R	none	CTAGACAGGAGAGAACAGGCCCCCATTGTTCAGTGCC
Competitor	137SOX9MUT1_F	none	CTAGGGCACTGCTCAATGGGGGCCTGTTCTCTCCTGT
Competitor	137SOX9MUT1_R	none	CTAGACAGGAGAGAACAGGCCCCCATTGAGCAGTGCC
Competitor	137SOX9MUT2_F	none	CTAGGGCACTGAACCCTGGGGGCCTGTTCTCTCCTGT
Competitor	137SOX9MUT2_R	none	CTAGACAGGAGAGAACAGGCCCCCAGGGTTCAGTGCC
Competitor	137SOX9MUT3_F	none	CTAGGGCACTGCTCCCTGGGGGCCTGTTCTCTCCTGT
Competitor	137SOX9MUT3_R	none	CTAGACAGGAGAGAACAGGCCCCCAGGGAGCAGTGCC

Table 2.7: Oligonucleotide sequences for EMSA testing of SOX9 interaction within -137kb enhancer. Oligonucleotides containing SOX9 consensus motif and flanking sequence were used to create probes labelled at the 5’with near-infrared dye, and with unlabelled oligonucleotides as competitors in EMSA. SOX9 consensus motif was mutated through substitution of CAA core motif nucleotides. Forward and reverse oligonucleotides were annealed for each sequence of interest

Probe/ competitor	Sequence name	label	Sequence 5’ to 3’
Probe	148sox9-WtP_F	DY682	CCATCTGTCAACAGAGGCTCTCAGCTGCTT
Probe	148sox9-WtP_R	DY682	AAGCAGCTGAGAGCCTCTGTTGACAGATGG
Competitor	148sox9-1WTcomp1_F	None	CCATCTGTCAACAGAGGCTCTCAGCTGCTT
Competitor	148sox9-1WTcomp_R	None	AAGCAGCTGAGAGCCTCTGTTGACAGATGG
Competitor	148sox9-1MutComp_F	None	CCATCTGTCCCCCGAGGCTCTCAGCTGCTT
Competitor	148sox9-1MutComp_R	None	AAGCAGCTGAGAGCCTCGGGGGACAGATGG
Probe	148sox9-2-WtP_F	DY682	GCCACATCTAATGTTTATGACCTAACGCCT
Probe	148sox-2-WtP_R	DY682	AGGCGTTAGGTCATAAACATTAGATGTGGC
Competitor	148sox9-2-WtC_F	None	GCCACATCTAATGTTTATGACCTAACGCCT
Competitor	148sox9-2-WtC_R	None	AGGCGTTAGGTCATAAACATTAGATGTGGC
Competitor	148sox9-2-MutC_F	None	GCCACATCTAATAGGTATGACCTAACGCCT
Competitor	148sox9-2-MutC_R	None	AGGCGTTAGGTCATACCTATTAGATGTGGC

Table 2.8: Oligonucleotide sequences for EMSA testing of SOX9 interaction within -148kb enhancer. Oligonucleotides containing SOX9 consensus motif and flanking sequence were used to create probes labelled at the 5’with near-infrared dye, and with unlabelled oligonucleotides as competitors in EMSA. SOX9 consensus motif was mutated through substitution of CAA core motif nucleotides. Forward and reverse oligonucleotides were annealed for each sequence of interest

Probe/ competitor	Sequence name	label	Sequence 5’ to 3’
Probe	-230soxIR_F	IRD700	CAGGGACAAAGTACCTTTGTCAACG
Probe	-230soxIR_R	IRD700	CGTTGACAAAGGTACTTTGTCCCTG
Probe	230soxIRMUT_F	IRD700	CAGGGGCGGGGTACCGGGGGCAAC
Probe	230soxIRMUT_R	IRD700	GTTGCCCCCGGTACCCCGCCCCTG
Competitor	230sox_WTC_F	none	CAGGGACAAAGTACCTTTGTCAACG
Competitor	230soxWTC_R	none	CGTTGACAAAGGTACTTTGTCCCTG
Competitor	230sox_MUTC1_F	none	CAGGGACGGGGTACCTTTGTCAACG
Competitor	230sox_MUT1C_R	none	CGTTGACAAAGGTACCCCGTCCCTG
Competitor	230sox_MUTC2_F	none	CAGGGACAAAGTACCGGGGTCAACG
Competitor	230sox_MUTC2_R	none	CGTTGACCCCGGTACTTTGTCCCTG
Competitor	230sox_MUTC3_F	none	CAGGGACGGGGTACCGGGGTCAACG
Competitor	230sox_MUTC3_R	none	CGTTGACCCCGGTACCCCGTCCCTG

Table 2.9: Oligonucleotide sequences for EMSA testing of SOX9 TFBS within -230kb enhancer. Oligonucleotides containing SOX9 consensus motif and flanking sequence were used to create probes labelled at the 5’with near-infrared dye, and with unlabelled oligonucleotides as competitors in EMSA. SOX9 consensus motif was mutated through substitution of CAA core motif nucleotides. Forward and reverse oligonucleotides were annealed for each sequence of interest

2.9.2.3 In vitro protein production

Proteins of interest were generated using an *in vitro* transcription and translation system with expression vectors containing protein of interest as template. The T7/SP6 TNT® Coupled Reticulocyte Lysate System (Promega, L5020) was used in accordance with the manufacturer's guidelines.

SOX9 protein was generated using SOX9-4xFLAG T7 plasmid as template, a gift from Veronique Lefebvre (Lefebvre *et al.* 1998). RUNX2 protein was generated using pcDNARunx2 expression vector.

2.9.2.4 Gel preparation

Native polyacrylamide gels were used with Tris-borate-EDTA (TBE) based buffer system. The Hoeffer SE400 electrophoresis system was used for gel running. Gel mixes were composed of 30.5mL distilled water, 4.25mL 5xTBE, 5.1mL 40% Bis-acrylamide solution (Bio-Rad, 161-0144), 140.8µL 20% (w/v) ammonium persulfate (Sigma Aldrich, 09913) solution and 40µL N, N, N', N'-Tetramethylethylenediamine (TEMED) (Sigma Aldrich, T928). Cast gels were left to set for one hour at room temperature before storage overnight at 4°C

2.9.2.5 Protein-oligonucleotide binding reactions

The Odyssey® EMSA Buffer kit (Li-Cor, 82907910) was used as the source for all reaction components and loading dye; aside from the protein, oligonucleotides and water.

Basic reactions consisted of:

Component	*Volume*
10X Binding buffer	2 µL
25mM Dithiothreitol (DTT)/ 2.5%Tween® 20	2µL
1µg/µL Poly(dI-dC)	0.5µL
100nM oligonucleotide probe	2µL
Protein	2µL
Nuclease free water	to 20µL

Binding reaction optimisation experiments were carried out for each combination of oligonucleotide probe and protein. This involved replicate binding reactions with an additional microlitre of additives supplied in the Buffer kit: 50% glycerol, 1% NP-40, 1M potassium chloride (KCl), 100mM (Magnesium chloride) $MgCl_2$ and 200mM EDTA. Comparisons of the effects of the additives on binding reactions were drawn upon examination of gel image, with the additive that yielded clearest shift used for subsequent reactions. Where competitor oligonucleotides were used, they were used in excess of either 25x or 50x the concentration of the oligonucleotide probe.

Once assembled, reactions were incubated in the dark for 20 minutes at room temperature prior to the addition of 2µL 10X Orange Loading Dye.

2.9.2.6 Gel running

Gels were pre-ran at 160V (10V/cm gel) for one hour at 4°C. After this period, samples were loaded and gels were run at 160V for three hours at 4°C in the dark.

2.9.2.7 Gel imaging

The Odyssey® CLx Infrared Imaging System in conjunction with Image Studio software (both Li-Cor) was used for gel visualisation. After gels had ran, the tank was quickly unassembled before the gel was shielded from light whilst still within glass plates. Gels remained in plates for imaging, which were wiped with isopropanol prior to imaging. Scans of gels were conducted using the following parameters; 700nm channel, focus offset of 4mm, 169µM resolution, intensity of eight and medium quality. Images were converted to black on white using the Image Studio Software.

3. Results

3.1 Prediction of Ccn2 enhancers and their positioning within a TAD

3.1.1 Identification of putative *Ccn2* enhancers

The region upstream of *Ccn2* was examined *in silico* in order to identify enhancers that may regulate transcription of the gene. Given the well-established role of *Ccn2* within cartilaginous tissue, datasets and annotation related to this tissue type was prioritised in the prediction of enhancers. Within the mm9 assembly of the Genome Browser, the sequence encoding *Ccn2* resides on the q arm of chromosome 10 at position chr10:24,315,532-24,317,417. The entire transcript from *Ccn2*; including UTR spans from chr10:24,315,248-24,318,488. An area of 300kb encompassing from the nearest gene upstream of *Ccn2*; monooxygenase DBH-like 1 (*Moxd1*), until approximately 2kb downstream of the *Ccn2* coding sequence was the focus of enhancer predictions (chr10:24,020,350-24,320,350), with a UCSC Genome Browser session created to examine this region, as illustrated in Figure 3.1.

Potential enhancers were identified using epigenetic annotations associated with enhancer function as outlined in Chapter 1.3 and Chapter 2.1 (datasets detailed in Chapter 5 appendices Table 5.1). These datasets were primarily gathered from the ENCODE consortium. Data concerning coding genes (Figure 3.1 B), histone posttranslational modification (Figure 3.1 C), DNase I hypersensitivity (Figure 3.1 D) and intra-species conservation (Figure 3.1 E) were compiled. The promoter region for *Ccn2* was identified using H3K4me3 signal at E14.5; with a single sharp peak in close proximity to the *Ccn2* gene (Figure 3.1 B and C). The promoter was also identified using lift-over of previously published human sequences examining this region (Leask *et al.* 2003). An E14.5 limb H3K4me1 signal track was used to identify potential enhancers, with several peaks across the region of interest (Figure 3.1 C). Limb based H3K27ac signal at E14.5 was also used to identify *cis*-regulatory elements, there was a sharp peak for this modification at the promoter region for *Ccn2* and further peaks across the region of interest. The distribution of this modification in comparison with H3K4me1 enabled preliminary prediction of enhancer activity. Chromatin state was further interrogated using DNase I hypersensitivity tracks from limb and mesoderm samples at E11.5 in addition to adult lung fibroblast at eight weeks. Peaks in each of these tracks were distributed throughout the region of interest. Highly conserved stretches of sequence were also observed across the region of interest, indicating important regulatory function.

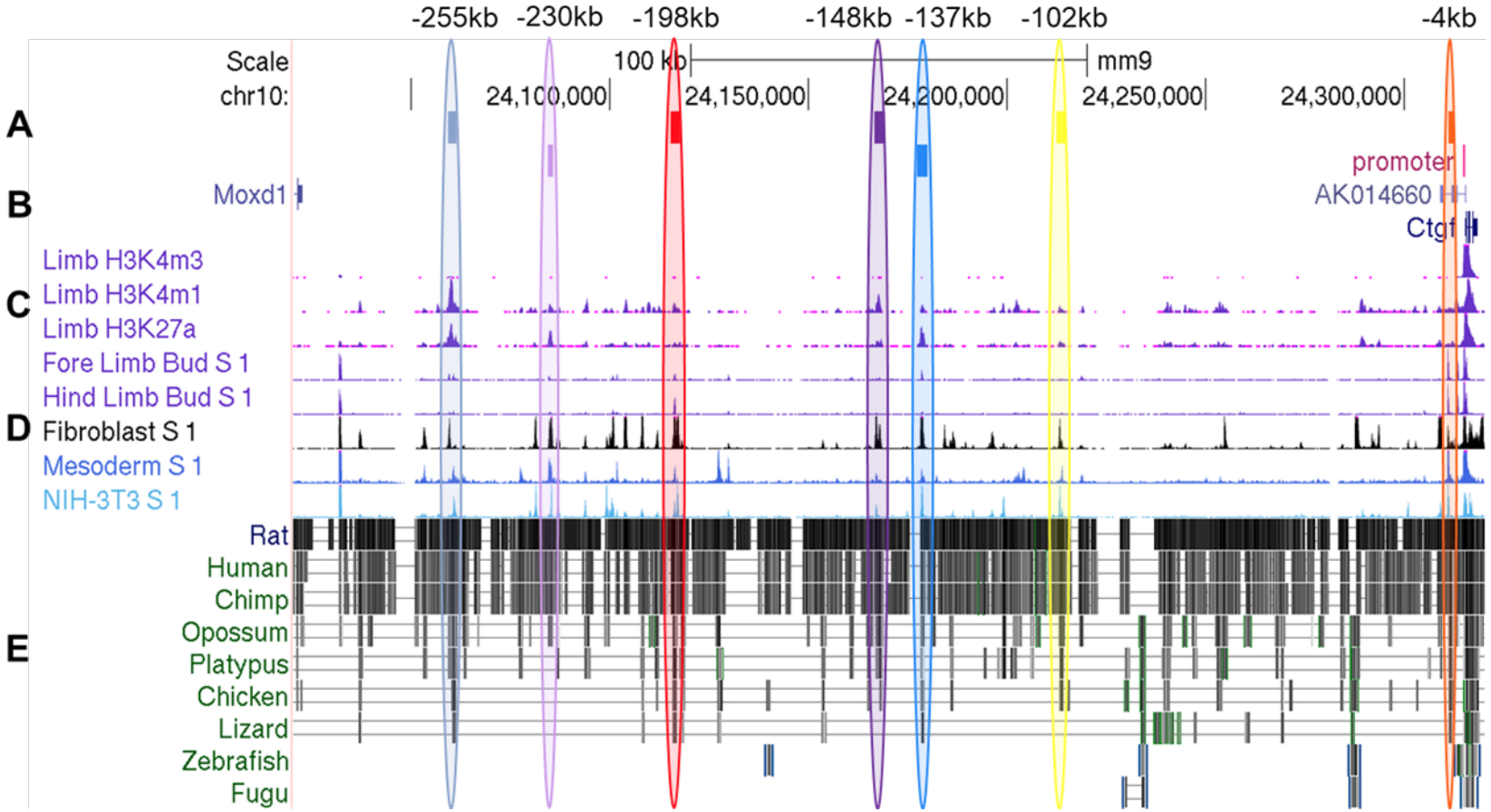


Figure 3.1: Identification of putative enhancer regions upstream of *Ccn2* (*Ctgf*). The non-coding genomic region between *Ccn2* (depicted as *Ctgf*) and nearest 5' gene *Moxd1* (B) served as the basic region in the identification of enhancers. Putative enhancers (A) and coloured loops) were identified on the basis of histone posttranslational modification, DNase I hypersensitivity (D) and evolutionary conservation of sequence (E). Enhancers were identified at -4kb, -102kb, -137kb,-148kb, -198kb, -230kb and -256kb from the *Ccn2* transcription start site.

Seven putative enhancers of *Ccn2* transcription within limb tissue were identified where peaks for these characteristics overlapped (Figure 3.1 coloured blocks A and loops) and named according to the position of the putative enhancer in kb relative to the transcription start site for *Ccn2* at position chr10:24,315,248 (Table 3.1).

Putative Enhancer	Genomic coordinates	Size (bp)
-4kb	chr10:24311246-24312541	1295
-102kb[10]	chr10:24212606-24214888	2282
-137kb[10]	ch10:24177497-24180069	2572
-148kb	chr10:24166687-24169261	2574
-198kb	chr10:24115344-24117739	2395
-230kb[10]	chr10:24084486-24085909	1423
-255kb[10]	chr10:24059327-24061377	2050

Table 3.1: Genomic coordinates of putative enhancer regions identified in ENCODE UCSC Browser. Enhancers were named according to distance from *Ccn2* TSS.

Unfortunately, there is not an H3K27me3 dataset available for the limb histone posttranslational modifications track used, so predictions of enhancer poising and inactivity are limited. However, basic predictions of activity can still be made on the basis of H3K4me1 and H3K27ac modification. For example, the peaks of H3K4me1 vs H3K27ac within -4kb, -102kb, and -255kb are similar, and in the -137kb and -230kb regions the H3K27ac peak is larger than H3K4me1. These observations suggest an active regulatory role for these regions in the limb at E14.5. This contrasts the -148kb region which has a greater peak for H3K4me1 than H3K27ac which indicates that this region may be in a poised state in this temporospatial context.

3.1.2 Prediction of *Ccn2* positioning within a TAD

There must be higher order organisation of the locus surrounding *Ccn2* in order to facilitate function of multiple enhancers in a cell type, or differential organisation of the enhancers in several tissue types. This therefore raises the question as to where *Ccn2* is positioned within a topologically associated domain. On the basis of previous studies, it would be assumed *Ccn2* lies within the same TAD as cognate

[10] Enhancers that have been reported in publication are also known by the distance from the 3' enhancer boundary to *Ccn2* TSS: -102kb is -100kb; -137kb is -135kb; -230kb is -229kb and -255kb is -254kb (Frost *et al.* 2018).

cis-acting regulatory regions (De Laat and Duboule 2013). A TAD containing *Ccn2* has not been specifically examined in the literature, but publicly available datasets can be used to predict the positioning of *Ccn2*, and enhancer regions described herein within a TAD.

Firstly, the aforementioned Hi-C data from Dixon *et al*. (2012) (Chapter 1.3) is available at as part of the '3D Genome Browser' http://promoter.bx.psu.edu/hi-c/view.php (Wang *et al.* 2018). Within mouse embryonic stem cells, within a 1.5 megabase (mb) viewpoint there is an enrichment of interactions between sequences spanning from upstream of *Ccn2* (*CTGF*) near neighbouring *Moxd1* to downstream of *CCN2*, near *Enpp1* (Figure 3.2). This region is annotated to be a TAD using this resource and dataset (Dixon *et al.* 2012; Wang *et al.* 2018) (highlighted blue box, Figure 3.2).

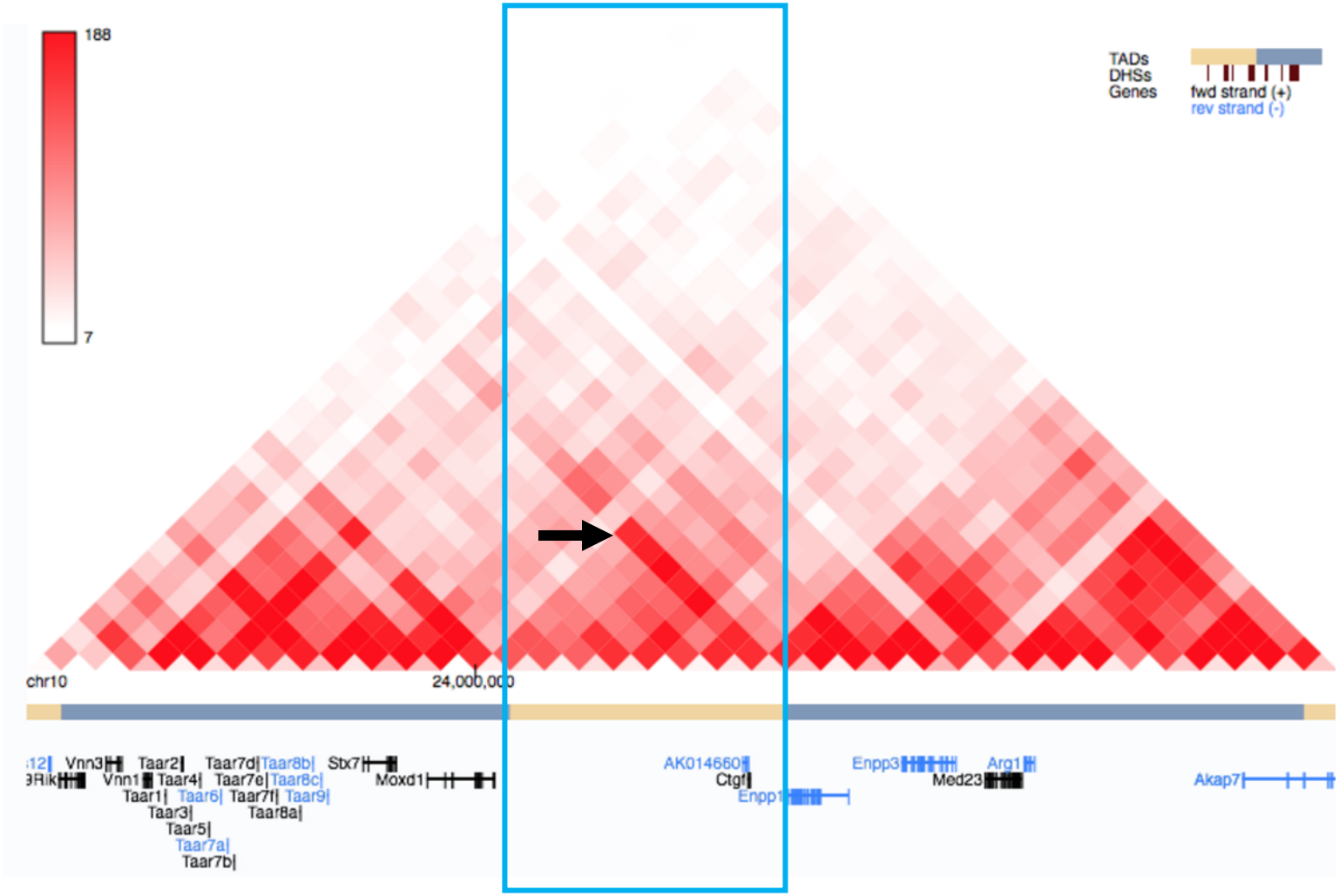


Figure 3.2: *CCN2* (*CTGF*) positioning within a TAD based on Hi-C interrogation of chromatin interactions. *Ccn2* (*Ctgf*) is predicted to lie within a TAD that predominantly flanks upstream of the gene. With enrichment interactions (red intensity) peaking in the middle of the putative TAD region (black arrow). The predicted TAD is highlighted with a blue box.

The 3D Genome Browser also contains Chromatin Interaction by Paired-End Tag Sequencing (ChIA-PET) datasets, which is a further technique used in the examination of chromatin interactions via protein, and ChIP-Seq datasets for CTCF and cohesin interactions (Wang *et al.* 2018). This is an important aspect of the prediction of TADs given that both of these proteins are involved in the looping and isolation of chromatin that constitutes TAD structure, as detailed in Chapter 1.3. (Dixon *et al.* 2016; Ong and Corces 2014; Rao *et al.* 2017). Data pertaining to these interactions within both human and mouse tissues can be examined, which was carried out for sequence in the vicinity of *CCN2* and *Ccn2* (Chapter 5.5). The regions of genomic sequence predicted to be looped through CTCF and cohesin can be compiled into a track on the UCSC Genome Browser with the 1.5mb viewpoint used in the Hi-C dataset (Figures 3.3 and Figure 3.2 respectively), alongside GEO-DataSet ChIP-Seq data for CTCF and cohesin (GSE55045 and GSM2169995 respectively, Figure 3.3 E and D respectively) (Barutcu *et al.* 2014; Kim *et al.* 2018). An ENCODE/LICR track for limb CTCF ChIP-Seq can also be compiled into the UCSC Genome Browser (Figure 3.3 E). There is enrichment for CTCF and cohesin interaction with the sequences in close vicinity to the ends of the Hi-C predicted TAD. Further datasets from the 3D-Genome Interaction Viewer and Database (Yang *et al.* 2018), available at http://kobic.kr/3div/ (Chapter 5.5), also allow visualisation of chromatin interactions and predictions of the localisation of *Ccn2* and *CCN2* within a TAD.

All of this evidence suggests that *Ccn2* sits within a TAD approximately 300kb long stretching 5' from approximately 8kb upstream of the -255kb putative enhancer to 3' approximately 35kb downstream of *Ccn2* (highlighted with red box in Figure 3.3). This therefore reinforces the findings of the current project, as enhancers are predominantly found within the same TAD as their target gene (Nuebler *et al.* 2018), and each of the enhancer regions described herein is predicted to be located in the same TAD as *CCN2*. Future work should seek to understand the higher order chromatin looping that constitutes this putative TAD and how this impacts on enhancer function. For example, further CTCF and cohesin sites may be utilised throughout the TAD in order to isolate enhancers that are not active and prevent interaction with the *Ccn2* promoter region, thereby reinforcing specific temporospatial function of each of the enhancers (Ren *et al.* 2017).

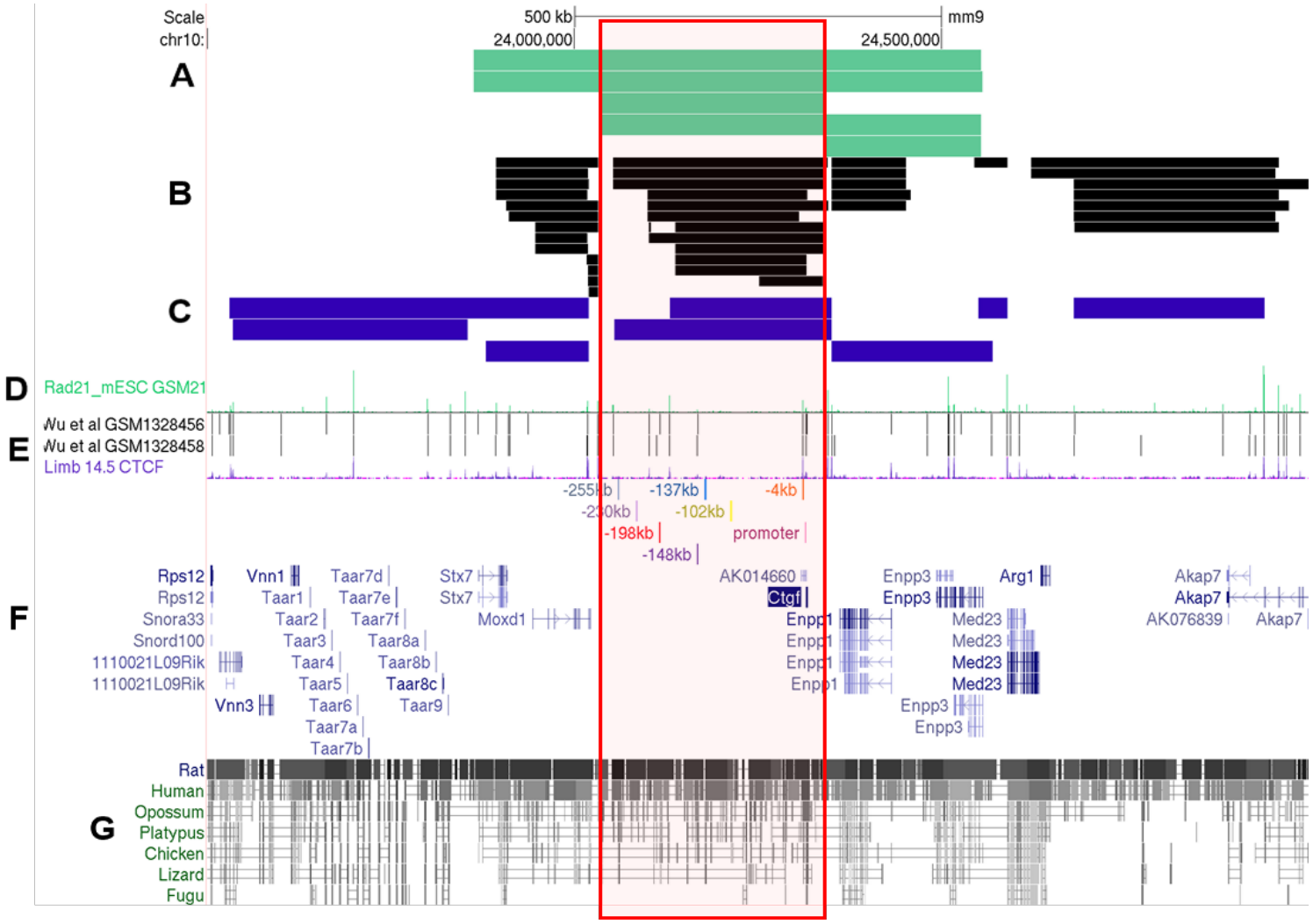


Figure 3.3: *In silico* prediction of *Ccn2* positioning within a TAD in the UCSC Genome Browser. Import of datasets pertaining to CTCF and cohesin binding (Barutcu *et al.* 2014; Kim *et al.* 2018; Yang *et al.* 2018), (A, B, C, D and E) allow prediction of chromatin organisation and TAD structure within a 1.5mb viewpoint encompassing *Ccn2* (*Ctgf*) and neighbouring genes (F). The putative TAD predicted in the Hi-C dataset and CTCF (A and B) and cohesin (C and D) interactions is highlighted with a red box. Conserved sequences (G) occur both within and between coding sequences.

Another interesting facet in the organisation of the enhancers is the prospect of super-enhancer function. As outlined in Chapter 1.3, super-enhancers are clusters of enhancers that function collaboratively under the control of cell lineage-specific TF leading to high levels of target gene transcription (Whyte *et al.* 2013). Overlap in the temporospatial activity of the enhancers described herein could therefore constitute super-enhancer function in ensuring robust expression of *Ccn2* in chondrocytes during embryonic development. Ohba *et al*. (2015) tried to define super-enhancers in chondrocytes using the ethos of Whyte *et al*. (2013), with SOX9 as the master transcription factor. On the basis of ChIP-Seq data validated SOX9 TFBS, this study suggested that there is a super-enhancer spanning approximately 8.2kb from upstream of the -4kb enhancer described herein, to approximately 1kb downstream of *Ccn2*. The genomic occupancy of Mediator within these regions would have to be determined in order to validate super-enhancer predictions, as this is a criteria upon which super-enhancer is judged (Pott and Lieb 2015). However, Ohba *et al*. (2015) also suggest that SOX9 binds to several cognate enhancer elements for genes involved in chondrocyte behaviour and function, without the super-enhancer annotation.

3.2 Multiple enhancers are active during embryonic development

Preliminary experiments were conducted in order to assess the capacity of the -4kb, -102kb, -137kb, -198kb, -230kb and -255kb putative enhancers to function as regulators of the expression of a *LacZ* reporter gene during embryonic development at E15.5 (Frost *et al.* 2018). Whilst the *in silico* datasets in Figure 3.1 concerned E14.5, the E15.5 time-point was used for *in vivo* experiments as cartilaginous tissue is more substantial at the latter time point, and endochondral ossification is more advanced (Kaufman 2003). X-gal staining, whole mount imaging and histological sectioning and staining of embryos at E15.5 was carried out by Dr Ian Li and Christoph Zimmer for each enhancer aside from -148kb short and -230kb region. More data was gathered for the function of the -148kb enhancer and so there are sub-chapters dedicated to this enhancer and a truncated version of it.

3.2.1 -4kb and -255kb putative enhancers do not seem to function at E15.5

For the -4kb enhancer, two attempts were made to establish transgenic lines that harboured the enhancer-Hsp68-*LacZ* construct by Dr Ian Li. Firstly, microinjection of construct (as outlined in Chapter 2.3 and 2.5) was carried out; from which a total

of 26 mice were born. Three of these were genotyped as being positive for the transgene. In pairings these founder (F_0) together, or pairings of their progeny (F_1) for E15.5 with wild-type mice, no X-gal positive staining was observed in any embryo at E15.5. For the second attempt, two mice were born, of which neither were genotyped to have either transgenic construct and so, enhancer function at E15.5 was not assessed. The capacity for the -255kb putative enhancer to drive *LacZ* expression was also examined at E15.5, there was negative staining for all transgenic E15.5 embryos generated.

3.2.2 -102kb is active within the vasculature at E15.5

A stable line of transgenic mice for examination of the enhancer located -102kb upstream of *Ccn2* was not created, instead assessment of the capacity for the enhancer to function at E15.5 was carried out with F_0 embryos by Dr Ian Ll. Transgene activity was observed in six embryos, as detailed in Table 3.2.

Founder	Whole mount staining intensity	Staining localisation	Intensity of local staining
1	++	Superficial vasculature, superficial eye orbit	+++
2	++	Superficial vasculature	+++
3	++++	Superficial vasculature	++++
4	++	Superficial vasculature	+++
5	++	Superficial vasculature	+++
6	+	Superficial vasculature	+

Table 3.2: Comparison of X-gal staining intensity and localisation across positive founder -102kbHsp68LacZ E15.5 embryos. Staining intensity (+ is negligible, ++ diffuse, +++ moderate, ++++strong, +++++intense) was recorded for a whole mount embryo, in addition to the specific location where staining was observed, enabling consistency of staining between F_0 to be assessed.

Staining occurred in close proximity to the superficial dermal layer each of these embryos, albeit to a varying degree. Three of the embryos exhibited staining in a highly similar manner as illustrated in Figure 3.4.

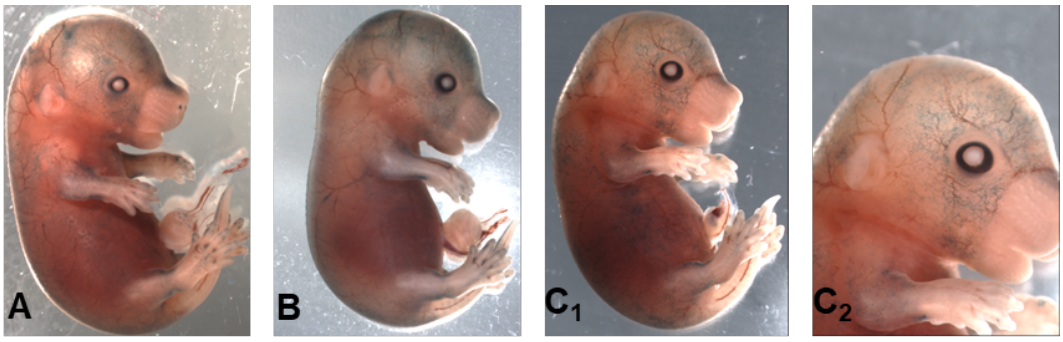


Figure 3.4: Representative whole mount images of X-gal stained E15.5 embryos containing the -102kbHsp68LacZGW reporter construct. Three of the embryos (A-C); corresponding to founders 2, 4 and 5 in Table 2, exhibited a highly similar pattern of punctate X-gal staining globally. Upon closer examination (C_2), blue staining co-localised with microvasculature, however larger blood vessels were not stained.

Histological sectioning of the embryos enabled more in-depth examination of the tissue in which the enhancer was functioning (Figure 3.5). Section images reinforced those from the whole mount imaging, with X-gal positive blue cells only observed within the superficial microvasculature. Tissue-specific staining was co-localised with endothelial cells of the capillary vessels of the dermal microvasculature. Moreover, erythrocytes were observed within branched blue vessel structures, lending credence to the notion that the enhancer was active within endothelial cells (Figure 3.5 B). β-galactosidase activity was absent in other populations of endothelial cells, such as within larger blood vessels, or the blood vessels within the growth plates of long bones undergoing endochondral ossification (Figure 3.5 C). X-gal staining was not observed within any other tissue; including the musculoskeletal system, as demonstrated in the cartilage anlage of the humerus (Figure 3.5 C), ribs (Figure 3.5 E), or tail vertebrae.

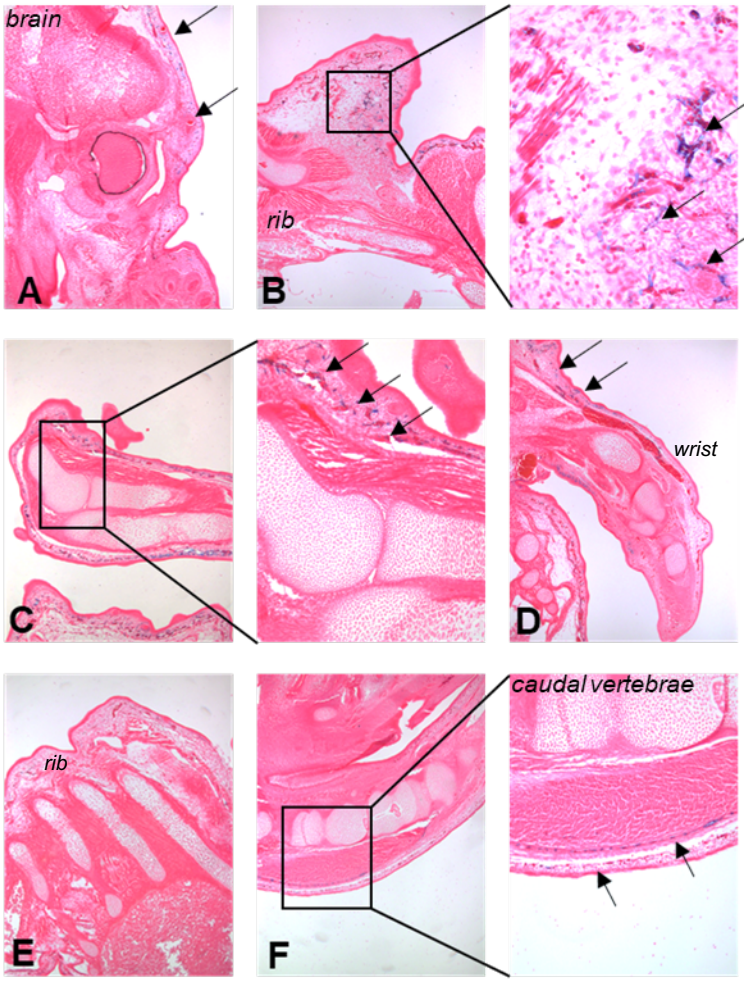


Figure 3.5: Histological examination of X-gal stained E15.5 -102kbHsp68LacZGW embryos. Blue staining, and therefore transgene activity was observed within superficial microvasculature of the dermis globally (highlighted black arrows). More specifically, staining was observed within the branched structures of the capillary vessels such as within the cranium (A) neck (B) and dermis surrounding the proximal, distal forelimb and caudal vertebrae (C, D and F respectively). There was no transgene activity in other vascular structures or blood vessels such as within the endochondral growth plates of the radius and ulna (C) or ribs (E).

3.2.3 -137kb functions within articular chondrocytes at E15.5

The capacity of the -137kb putative enhancer was examined in F_0 E15.5 embryos harbouring *LacZ* driven by the -137kb region (-137kbHsp68LacZ). Five founders were generated, with intensity and localisation of X-gal staining recorded in Table 3.3.

Founder	Whole mount staining intensity	Staining localisation	Intensity of local staining
1	++	Cranium, elbow, wrist, knee, ankle	++++
2	++	Elbow, wrist, knee, ankle	+++
3	++	Elbow, wrist, knee, ankle	+++
4	+	Wrist, ankle	++
5	+	Wrist, distal phalanx of paw	+

Table 3.3: X-gal staining localisation and intensity in E15.5 transgenic embryos containing the -137kbHsp68LacZ construct. Staining intensity (+ is negligible, ++ diffuse, +++ moderate, ++++strong, +++++intense). A highly similar pattern of staining was observed across 3 of the embryos with moderate staining of some joints.

Reproducible X-gal staining was observed in the wrist in each of the five F_0; albeit to varying extents. Tissue of the ankle was stained in four of the F_0. Two of the founders exhibited highly similar intensity and localisation of staining; as shown in Figure 3.6.

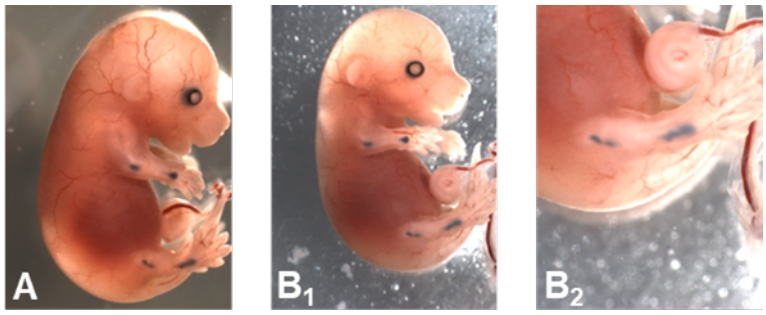


Figure 3.6: Representative whole mount imaging of β-galactosidase activity in E15.5 embryos harbouring the -137kbHsp68LacZ transgenic construct. X-gal staining was reproducibly confined to the articular regions of the elbow, wrist, knee and ankle (A and B).

Sub-populations of chondrocytes such as in the chondrocranium (Figure 3.7 A), rib (Figure 3.7 B) or within the broader endochondral skeleton precursor; such as within the diaphysis of the humerus, radius and ulna (Figure 3.7 C) did not exhibit X-gal staining. Where staining was observed in four of the founders it occurred in a lateral area of the articular joint, with a fifth founder demonstrating more diffuse staining across the joints Staining was not observed in the shoulder or hip (not shown) or intra-digit articular joints such as those in the manus (Figure 3.7 D). Upon histological sectioning, these observations were confirmed with positive staining

only occurring in the articular chondrocytes of the elbow (Figure 3.7 C), wrist (Figure 3.7 D) and ankle (Figure 3.7 E).

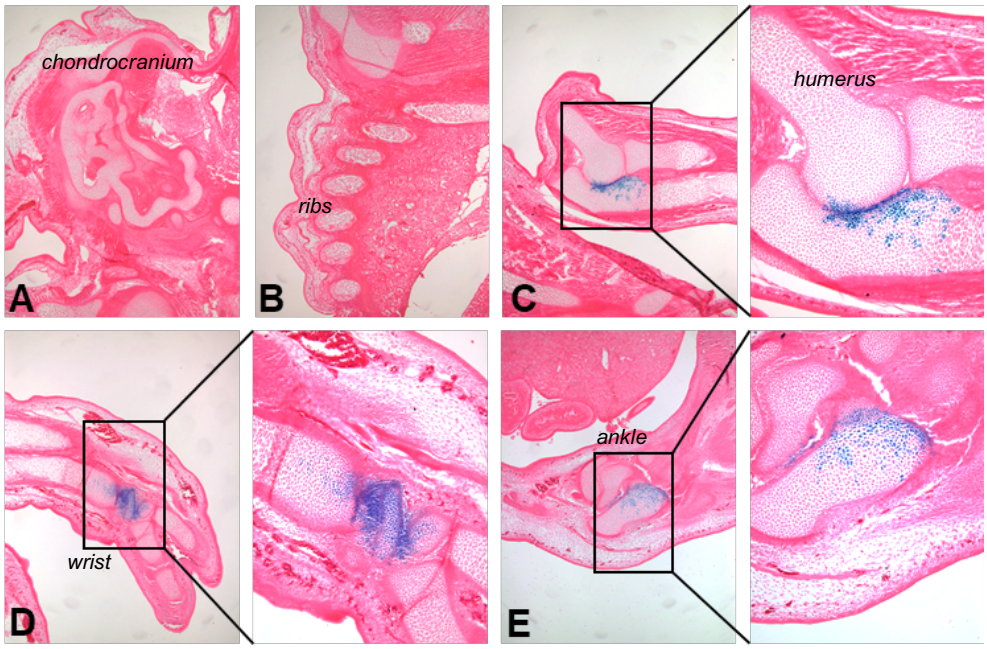


Figure 3.7: Representative histological sections of -137kbHsp68LacZ transgenic E15.5 embryos. Staining was not observed in chondrocranium primordium of the temporal bone (A), ribs (B), or precursor diaphysis region of the long bones such as the radius (C). Positive blue stained cells were most abundant within the vicinity of the articular surface of the radius (C), wrist (D) and ankle (E).

Where X-gal staining was present, it was observed to be stronger at articular joint surfaces. This occurred most strikingly in the elbow (Figure 3.7 C), where darker blue staining was observed proximal to the articular space, with stratification of staining intensity in cells located more distally from the articular surface.

An attempt was made to establish a line of -137kbshortHsp68LacZ transgenic mice, harbouring a conserved fragment of the -137kb enhancer (-137kb short). This shorter region was a truncated version of the -137kb enhancer containing approximately 1275bp of the middle of the sequence that was most highly conserved. The coordinates of this region are chr10:24,178,225-24,179,499 in the mm9 genome build. This region was also used in *in vitro* experiments. Three founders were genotyped as being positive for the transgenic construct but there was no transgene expression in E15.5 derived from any of these founders.

3.2.4 -198kb is active in multiple tissues at E15.5

The -198kb region was examined at E15.5 in three F_0 embryos by Dr Ian Li and Christoph Zimmer. Each founder that was positive in X-gal staining exhibited a different pattern of X-gal staining, as detailed in Table 3.4. Whole mount images were taken of each embryo, as shown in Figure 3.8.

Founder	Whole mount staining intensity	Staining localisation	Intensity of local staining
A	++++++	Subdermal globally, absent in distal phalanx	+++++
B	+++	Punctate superficial staining globally. Nares,	+++
C	++	Nasal bone primordium, nares	++++

Table 3.4: X-gal staining and intensity in E15.5 transgenic embryos containing the -198kbHsp68LacZ construct. Staining intensity (+ is negligible, ++ diffuse, +++ moderate, ++++strong, +++++intense) Intensity of staining varied between each embryo. Staining intensity and localisation varied between each of the 3 founder embryos.

One of the embryos stained very strongly globally (A, Table 3.4 and Figure 3.8 A). This contrasted the other positively stained embryos, in which blue staining was observed in a more diffuse manner.

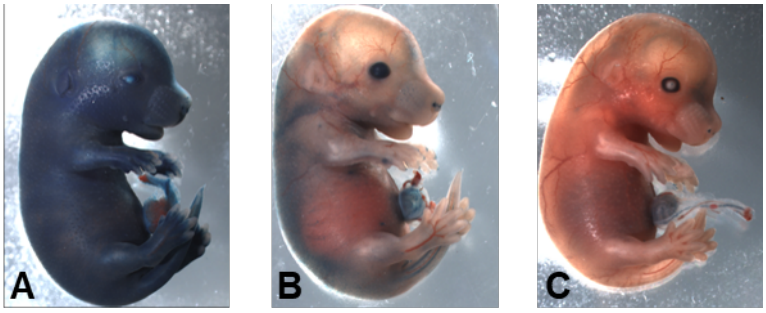


Figure 3.8: Whole mount imaging of X-gal stained -198kbHsp68LacZ embryos at E15.5. Three founder embryos exhibited blue staining of tissues. Transgene activity occurred in several tissues in each embryo.

Histological sectioning revealed that the intense staining observed in A occurred throughout the sagittal plane of the embryo. Staining was observed in several tissues and cell types. Staining was observed in the mesenchymal tissue of craniofacial dermal layer (Figure 3.9 A and B), hypertrophic chondrocytes within the humerus (Figure 3.9 C), articular chondrocytes of the elbow (Figure 3.9 D), mesenchymal tissue of the dorsal dermal layer (Figure 3.9 E) and chondrocytes of the ribs (Figure 3.9 F), articular chondrocytes of the hip (Figure 3.9 G), hypertrophic

chondrocytes within the hind limb (Figure 3.9 H) and fibroblastic tissue of the ligament/tendon attachment sites in the hind paw (Figure 3.9 I)

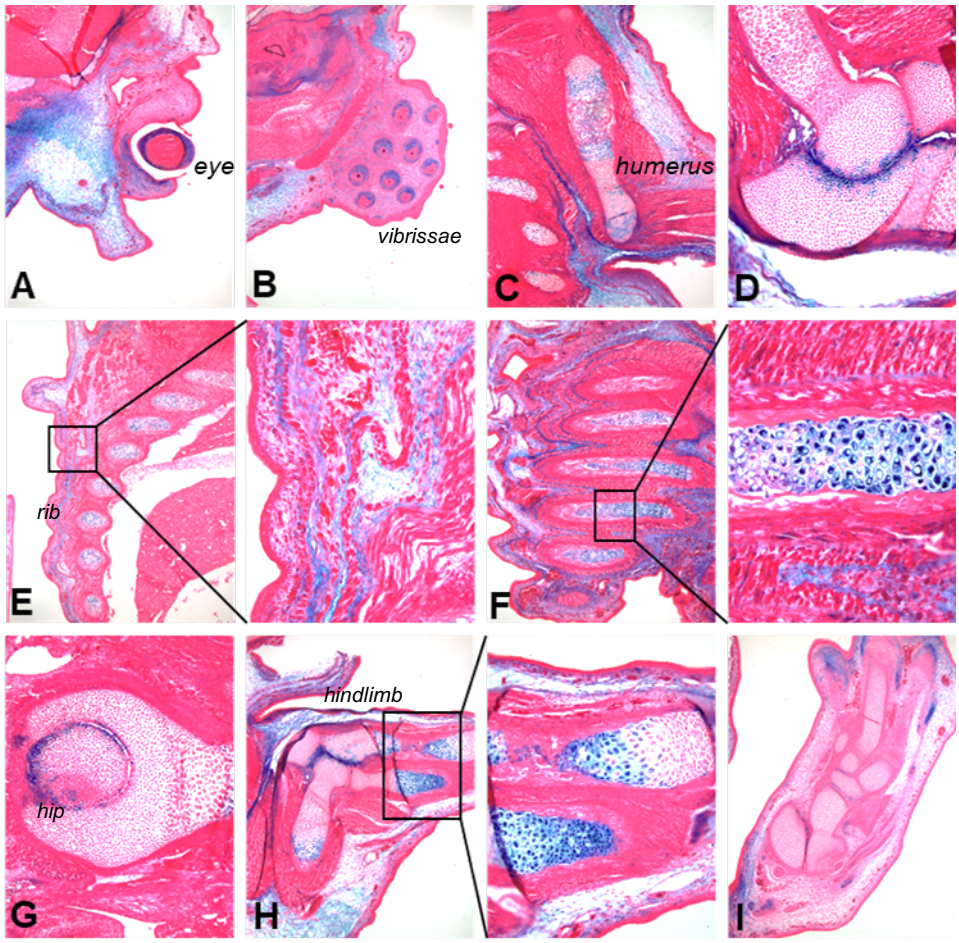


Figure 3.9: Histological section images of X-gal stained-198Hsp68LacZ E15.5 founder A. Strong X-gal staining was observed in a variety of tissues and cell types. Intense staining was observed within the reticular layer of the dermis, as demonstrated in the craniofacial (A) and dorsal (E) regions. The tissue surrounding the vibrissae follicles (B) also stained within the cranium. Staining occurred in several chondrocyte subpopulations including the hypertrophic cells of the humerus (C), ribs (F) in addition to the tibia and fibula (H). Articular chondrocytes of the elbow (D) and hip (G) also exhibited blue staining. Fibroblastic activity was also observed within ligamentous tissue, such as within the hind paw (I).

Strong X-gal staining was also observed within sub-populations of chondrocytes; however not all chondrocytes were X-gal stained. Staining within tissue undergoing endochondral ossification was stratified in line with the differentiation state of the chondrocytes. This was clearly illustrated within the humerus (Figure 3.9 C) and the tibia and fibula of the hind limb (Figure 3.9 F). Proliferative chondrocytes of the resting zone were negative for staining, whereas hypertrophic chondrocytes in proximity to the primary ossification centre exhibited strong X-gal staining. Osteoblastic cells within the ossification centre were not stained; therefore within endochondral tissue, *lacZ* expression was restricted to the hypertrophic

chondrocytes. Transgene expression was also observed within the articular chondrocytes of the elbow, hip and knee (Figures 3.9 D, G and H). Chondrocytes within the rib (Figure 3.9 E and F) also demonstrated strong X-gal staining.

The staining pattern observed within embryo A was not observed in histological sections of embryo B (Figure 3.10 A-C) or embryo C (Figure 3.10 D-G). Neither of these embryos demonstrated clear chondrocyte based X-gal staining. There was negligible staining within endochondral tissue within embryo C with single blue chondrocytes in a field of view (black arrow, Figure 3.10 E). Neither embryo displayed clear staining of any given cell type or tissue on a global scale.

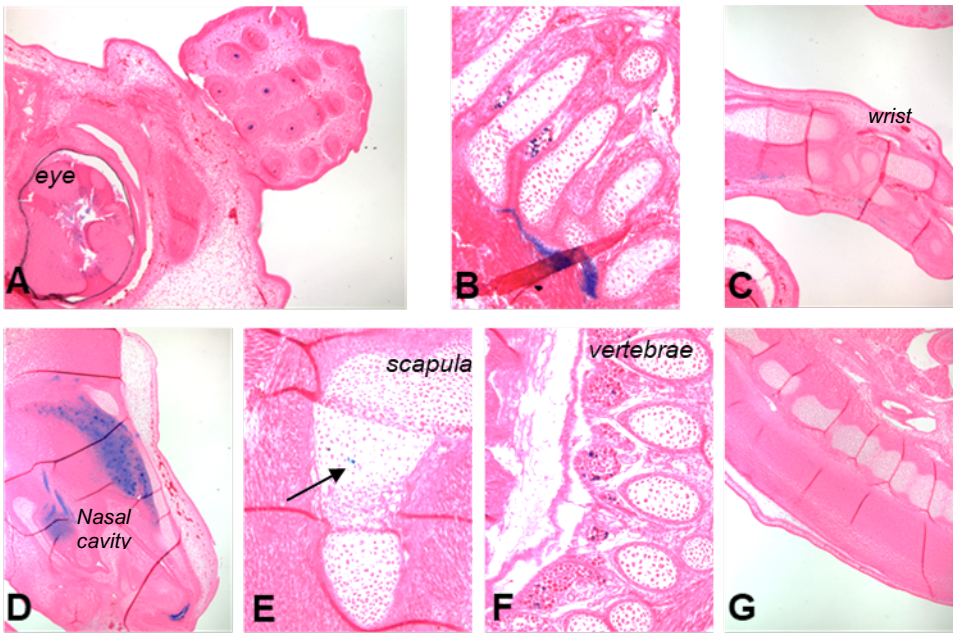


Figure 3.10: Representative histological samples from X-gal stained-198Hsp68LacZ E15.5 B and C. Sparse positive X-gal staining was observed in embryo B (A-C) with representative sections from the cranium (A), rib (B) and distal forelimb (C). Weak staining was also observed within embryo C such as near the nasal cavity (D), scapula (E), cervical vertebrae (F) and caudal vertebrae (G).

Within embryo B, the strongest X-gal staining occurred in line with ligamentous and tendinous tissue. The strongest staining in embryo C was observed within the nasal craniofacial region, whole mount staining suggested that this could have occurred in the primordial nasal cartilage, however sectioning revealed that the X-gal stained cells did not have chondrocyte morphology, with the chondrocytes of the primordial nasal septum negative for staining (Figure 3.10 D). Staining in this tissue was more closely allied with the localisation of the serous glands of nasal cavity.

3.2.5 -230kb functions in osseous tissue in development and adulthood

The putative enhancer located -230kb upstream of *Ccn2* was also examined using transgenic mice harbouring the enhancer driving expression of *lacZ* via an Hsp68 promoter (-230Hsp68LacZ). Rather than transient assessment of function in F_0, founders were left to develop to adulthood, with assessment of E15.5 transgene activity in F_1 or F_2. The activity of the -230Hsp68LacZ construct at E15.5 was assessed in lines of four founders which were genotyped as containing the transgene, as detailed in Table 3.5. Tissue samples were not available to assess enhancer function in adulthood for Founder A.

Founder	Whole mount staining intensity	Staining localisation	Intensity of local staining
A	+++++	Tissue subject to ossification- Frontal cranium, limbs; not as strongly in paws, vertebrate, ribs, tail	+++++
B	+++	Cranium; frontal bone region, nasal septum, inter-parietal and ex-occipital region. Spine. Sparse staining within paws	Cranium ++++ paws ++
C	-	Negative for staining	-
D	-	Negative for staining	-

Table 3.5: X-gal staining localisation and intensity in E15.5 transgenic embryos containing the -230kbHsp68LacZ construct. Staining intensity (+ is negligible, ++ diffuse, +++ moderate, ++++strong, +++++intense). There was no staining in any E15.5 embryos from founders C and D. Cranial staining was observed in all X-gal positive embryos.

Whole mount images of positively stained embryos are depicted in Figure 3.11.

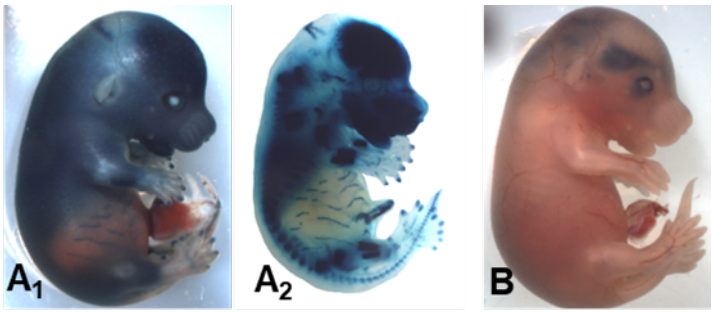


Figure 3.11: Whole mount imaging of X-gal stained -230kbHsp68LacZ E15.5 embryos. E15.5 originating from founder A (A) exhibited intense whole mount staining, soft tissue clearing (A_2) revealed strong staining of skeletal elements such as the cranium, ribs and limb. The staining of E15.5 from founder B was strongest in the cranium (B).

From the whole mount imaging of embryos derived from founder A, it was assumed that stained had occurred throughout the tissues which were X-gal positive. This was disproved upon histological sectioning with the observation that staining of skeletal elements was confined to the periosteum in regions including the orbital plate of the frontal bone (Figure 3.12 A), rib (Figure 3.12 C), and tibia and fibula of the hind limb (Figure 3.12 F).

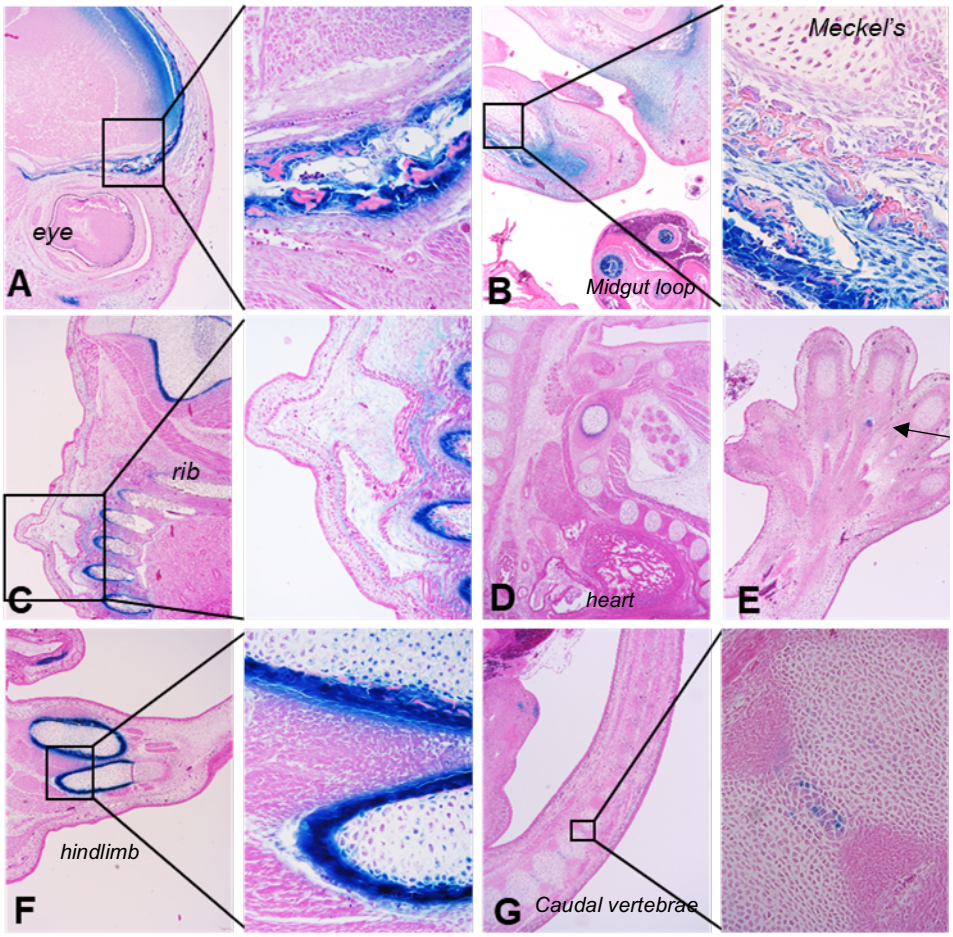


Figure 3.12: Representative histological section images of X-gal stained E15.5 progeny of -230kbHsp68LacZ founder *A*. Potent β-galactosidase activity was observed within osseous tissue. Periosteal tissue exhibited the greatest intensity of staining. Strong staining was observed in the frontal bone (A), mandible (B) and ribs (C). The costal cartilage (D) was devoid of staining, as was the endochondral cartilage of the digits (E). Further potent expression was observed within the periosteum of the tibia and fibula (F), contrasting sparse staining within the intervertebral disc (G).

The costal cartilage was negative for staining, with blue positive staining confined to the portion of the rib that undergoes ossification. This was also illustrated in the mandible (Figure 3.12 B) with intense staining of osteoblastic cells, surrounding the Meckel's cartilage which was negative for X-gal activity. *LacZ* activity was observed in primitive skeletal tissues undergoing both intramembranous and endochondral ossification processes. The chondrocytes of the endochondral anlage such as in

the tibia and fibula of the hind limb (Figure 3.12 F) were negative for X-gal activity further reinforcing the osteoblastic based expression of the transgene in this tissue. However, there was some limited staining of cartilaginous tissue such as within the intervertebral disc of the caudal vertebrae (Figure 3.12 G), in addition to sparse staining within the chondrocranium. Staining was also observed within fibroblastic cells of the reticular layer of the dermis (Figure 3.12 C) and at tendon and ligament attachment sites, such as in the paw (arrowhead, Figure 3.12 E). Soft tissue cleared embryo (Figure 3.11 A_2) demonstrated staining of the umbilicus region, sectioning revealed that this originated from the lumen of the mid-gut loop (Figure 3.12 B).

This expression pattern was not recapitulated in the E15.5 embryos stemming from founder B (Figure 3.13).

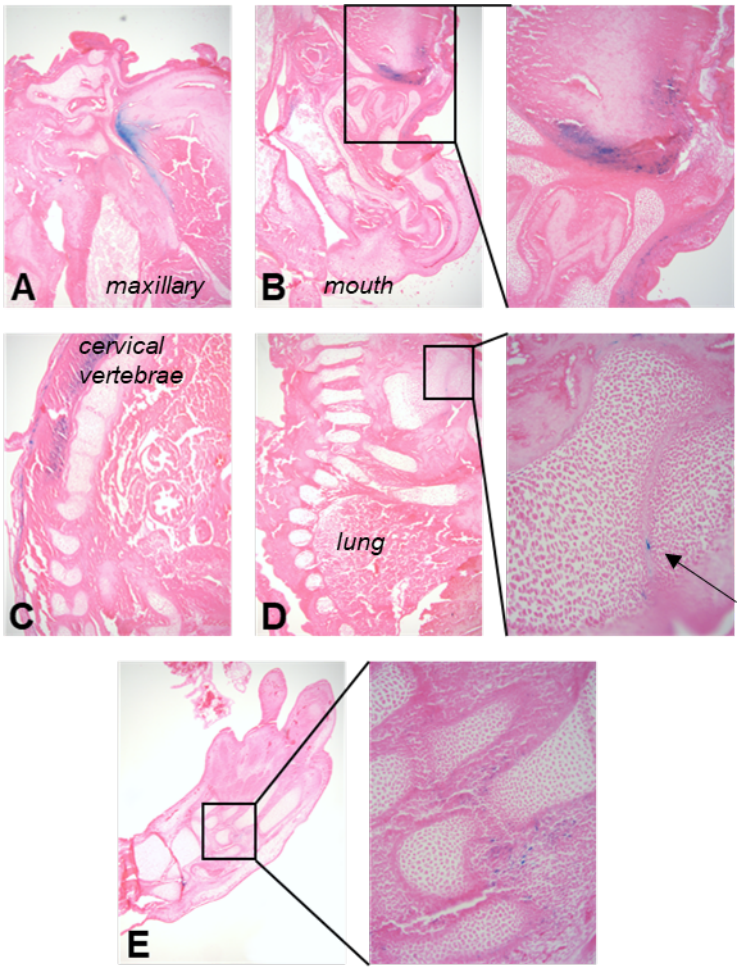


Figure 3.13*:* Representative histological section images of X-gal stained E15.5 progeny of 230Hsp68LacZ founder B. Few regions of blue X-gal stained cells were observed, predominantly occurring in the brain (A and B) in addition to the nervous tissue within the spine (C). Weak staining occurred within musculoskeletal tissues such as negligible chondrocytes staining (black arrow, D) and within the paw (E). There was no staining within osseous tissue.

Whilst the progeny from founder C did not exhibit transgene activity at E15.5, X-gal staining was observed postnatally. Firstly, at postnatal day seven (P7) β-galactosidase activity occurred within skeletal tissue alone. This staining was not in an intense manner, and was present within discrete areas of osseous tissue. Soft tissue clearing enabled this to be visualised more efficiently whole mount (Figure 3.14)

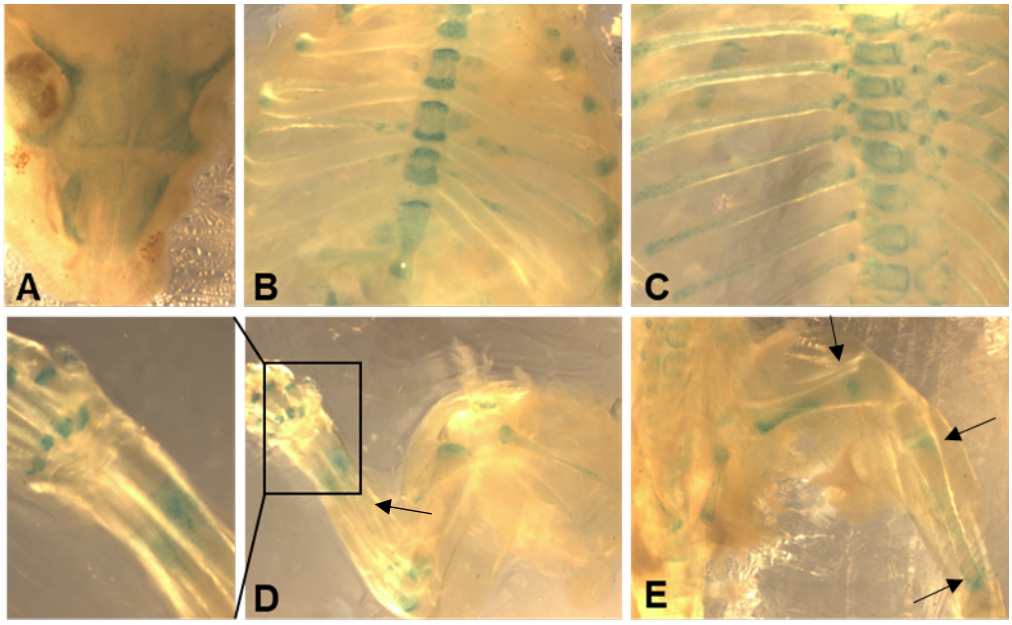


Figure 3.14: Whole mount imaging of soft tissue cleared X-gal stained progeny of P7 -230Hsp68lacZ founder C. X-gal staining only occurred in osseous tissues such as the cranium (A), sternebrae (B), thoracic vertebrae and dorsal portion of the ribs (C), forelimb (D) , in addition to the pelvis and hind limb (E). Within these tissues, staining was discrete yet with heterogeneous distribution, with strongest staining in the vicinity of the ends of the bones in a band-like pattern such as within the femur and tibia (arrows, E).

Within the skull, punctate staining was observed in the temporal region, contrasting the more intense blue staining of the frontal, nasal, maxillary and mandible bones (Figure 3.14 A). Within other osseous tissues, staining occurred proximal to the epiphysis, with weaker staining towards the diaphysis where there was negligible *LacZ* activity, for example within the radius and ulna (arrow Figure 3.14 D). Staining was not observed within cartilaginous tissues such as the xiphoid, articular joint surfaces of the elbow (Fig. 3.14D), knee (Fig. 3.14E) or hip (Fig. 3.14E). The specificity of staining to osseous tissue as oppose to cartilaginous tissue was exemplified in ribs (B) where the costal cartilage was negative for staining, whereas osseous tissues of the sternebrae and dorsal portion of the ribs (Fig. 3.14C) were positive. Histological sectioning allowed further refinement of the cells and tissues in which the transgene was active (Figure 3.15).

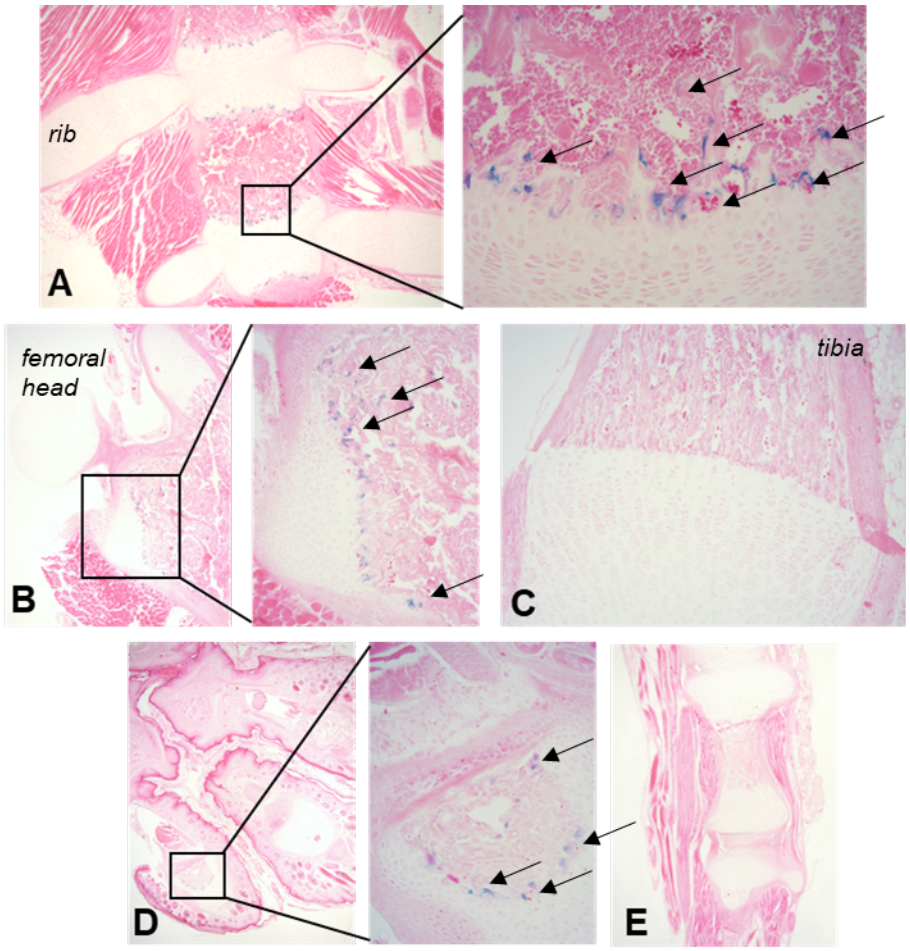


Figure 3.15: Representative histological sectioning images from P7 progeny of founder C of the -230Hsp68LacZ transgenic line. There was no transgene activity in any cartilaginous tissues such as the costal cartilage (A), or articular cartilage of the hip (B) or tail (E). Chondrocytes of growth plates were also negative for staining such as within the distal tibia (C). However, blue X-gal positive cells were observed in metaphyseal osseous tissue in close proximity to endochondral ossification fronts (black arrows) globally such as within the sternebrae (A), lesser trochanter of the femur (B) and phalanx (D).

Transgene activity in F_1 of founder C was also detected at six weeks of age, as shown in whole mount imaging in Figure 3.16. Six weeks was used as a time point due to fiscal limitations due to the costs associated with propagating the line for longer.

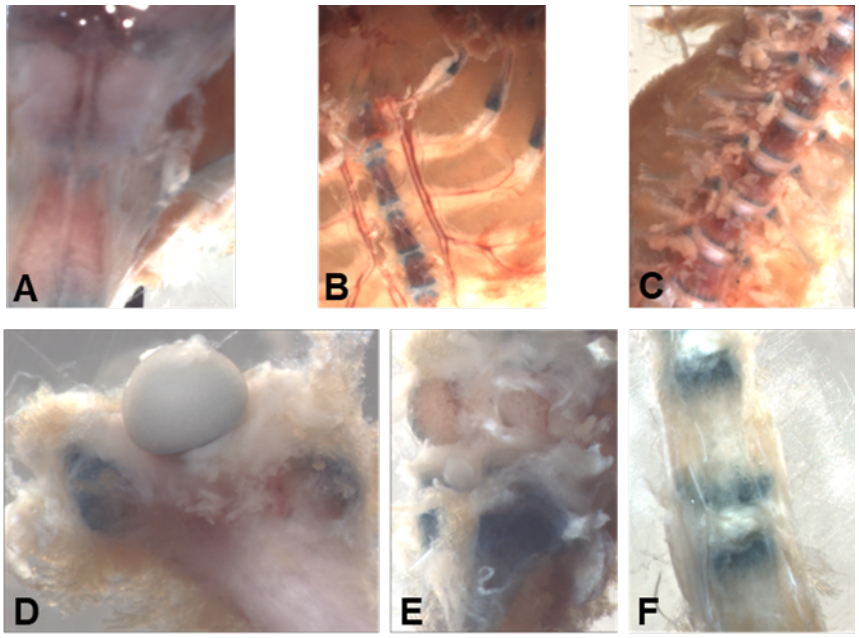


Figure 3.16: Whole mount images of X-gal stained six week old progeny from founder C of the -230Hsp68LacZ transgenic mouse line. Positive blue staining was observed within many skeletal elements such as; the cranium (A), sternebrae and ribs (B), vertebrae (C), femoral trochanter (D), tibia (E) and tail (F). More specifically, staining occurred within the metaphyseal and epiphyseal regions of bones.

As at P7, staining of tissue from mice aged six weeks occurred in osseous tissue and was observed in both whole mount staining and histological sectioning of tissues (Figure 3.16 and 3.17 respectively). This contrasted wild-type littermates that were negative for staining (data not shown). Within the cranium, staining was not observed globally with punctate staining within the nasal bone and maxillary (Figure 3.16 A). Cartilaginous elements such as the costal cartilage (Figure 3.16 B, Figure 3.17 C), intervertebral discs (Figure 3.16 C and F, Figure 3.17 B and D), and articular cartilage of the femoral head (Figure 3.16 D) and femoral condyles (Figure 3.16 E) were negative for β-galactosidase activity. However, X-gal staining was observed in close proximity to the cartilage of the epiphyseal plates, for example at the greater trochanter of the femur (Figure 3.16 D) and tibial plateau (Figure 3.16 F, Figure 3.17 C). The intensity of staining was reduced as proximity to the growth plate decreased; clearly demonstrated in the tibia (Figure 3.16 E, Figure 3.17 C), with dark blue staining observed at the metaphysis and loss of staining towards the diaphysis.

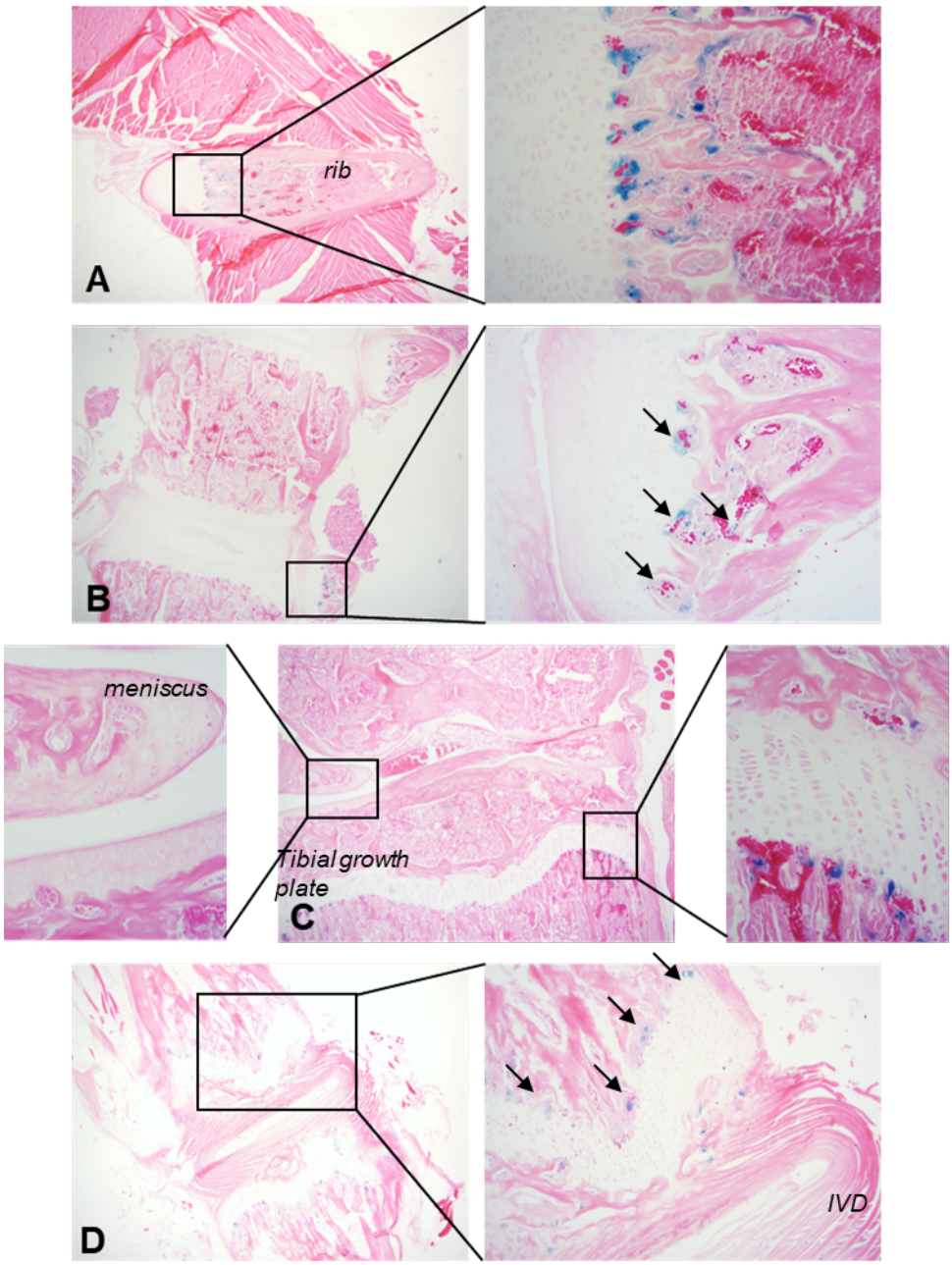


Figure 3.17: Histological section images of x-gal stained six week old offspring of founder C of the -230Hsp68LacZ transgenic mouse line. Chondrocytes were negative for any positive staining such as within the costal cartilage (A), thoracic vertebrae (B), articular cartilage of the knee (C) and intervertebral disc (IVD) (D). Transgene activity was observed within osteoblasts the primary spongiosa in close proximity to epiphyseal plates within the rib (A), proximal tibial growth plate (C) and tail (D).

LacZ expression was assessed in F_1 offspring of founder C at four months, and in founder C at six months of age. There was no X-gal staining of any tissue at these time points, as represented in whole mount images of founder C at six months (Figure 3.18). Regions that had stained at six weeks during postnatal bone growth in F_1 progeny were negative for transgene activity at this later time point; such as

in the rib (Figure 3.18 A), thoracic vertebrae (Figure 3.18 B), or caudal vertebrae (Figure 3.18 E).

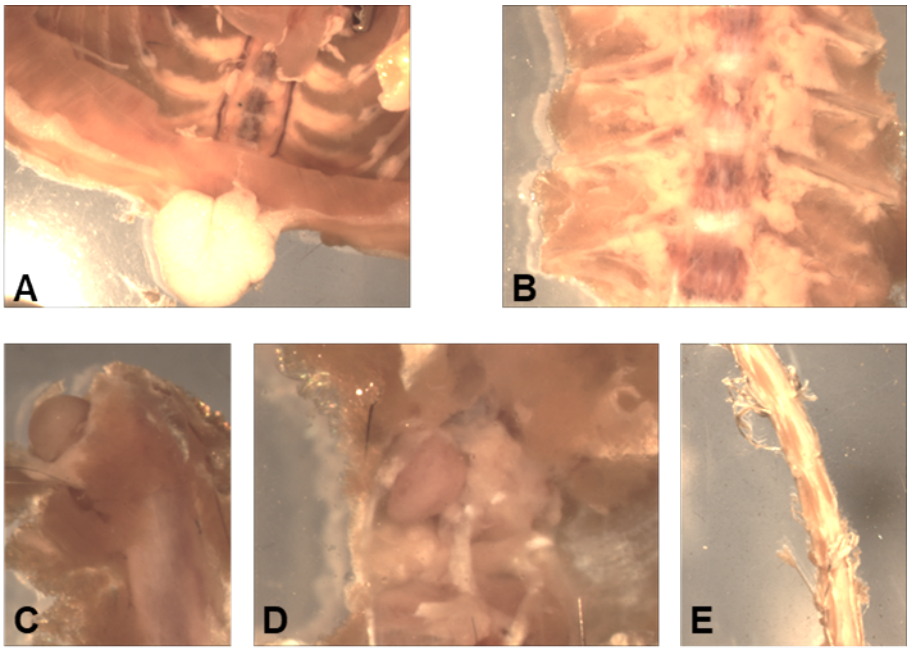


Figure 3.18: Whole mount images of X-gal stained -230Hsp68lacZ founder C aged six months. Blue staining was not observed in any tissue including the musculoskeletal system, as demonstrated in the ribs (A), thoracic vertebrae (B), hip (C) knee (D) and tail (E).

A Summary of the patterns in X-gal staining results for -230kb line variants is detailed in Table 3.6.

Founder	Time-point	Staining localisation
A	E15.5	Tissue subject to ossification- Frontal cranium, limbs; not as strongly in paws, vertebrate, ribs, tail
B	E15.5	Cranium; frontal bone region, nasal septum, inter-parietal and ex-occipital region. Spine. Sparse staining within paws
C	E15.5	Negative for staining
	P7	Osseous tissue; epiphysis of tissue that develops through endochondral ossification
	Six weeks	Osseous tissue; epiphysis of tissue that develops through endochondral ossification
	Four months	Negative for staining
D	E15.5	Negative for staining

Table 3.6*:* Table 3.6: summary of *LacZ* expression patterns driven by the -230kb enhancer across four variants of the -230Hsp68lacZ line. Expression within osseous tissue was common between founders A and C, albeit the prevalence of expression varied between the lines, with expression in A being strong at E15.5, whereas for the line derived from founder C, expression was greatest in the early postnatal period.

3.3 -148kb functions in founder embryos

The putative enhancer located -148kb upstream of *Ccn2* (Figure 3.1) exhibited strong function within chondrocytes and so function of this enhancer was scrutinised in greater detail than for the other enhancer regions. Assay of β-galactosidase expression was carried out in both founder and stable line (CTGF148) mice at several time points. The findings for this enhancer are arranged in order of developmental chronology, but the E15.5 time-point was characterised first, which led to the decision to further scrutinise a wider range of time-points. Unfortunately this process could not be carried out for the other enhancers owing to fiscal and time constraints.

3.3.1 -148kb functions at E11.5 in founders

Assessment of the -148kb enhancer to drive the expression of *lacZ* at E11.5 was carried out using X-gal staining. E11.5 was chosen as it precedes chondrogenesis, and would therefore aid assessment of chondrocyte based enhancer function, and the timing involved in this. A total of seven embryos were positively X-gal stained. A total of nine embryos were found to have the construct within their genome upon genotyping of placental samples. Transgene activity varied in each positive embryo, as detailed in the whole mount images of Figure 3.19. Some of the embryos gathered in this process (Fig. 3.19B, E and G) exhibited developmental delay and resembled embryonic day 8.5-9.5 (E8.5-E9.5) time-point more closely than E11.5. This made histological sectioning and comparisons of tissue specificity in transgene activity more difficult. The strong transgene activity within embryo C was reaffirmed upon histological sectioning, with staining of somitic tissue (Figure 3.20).

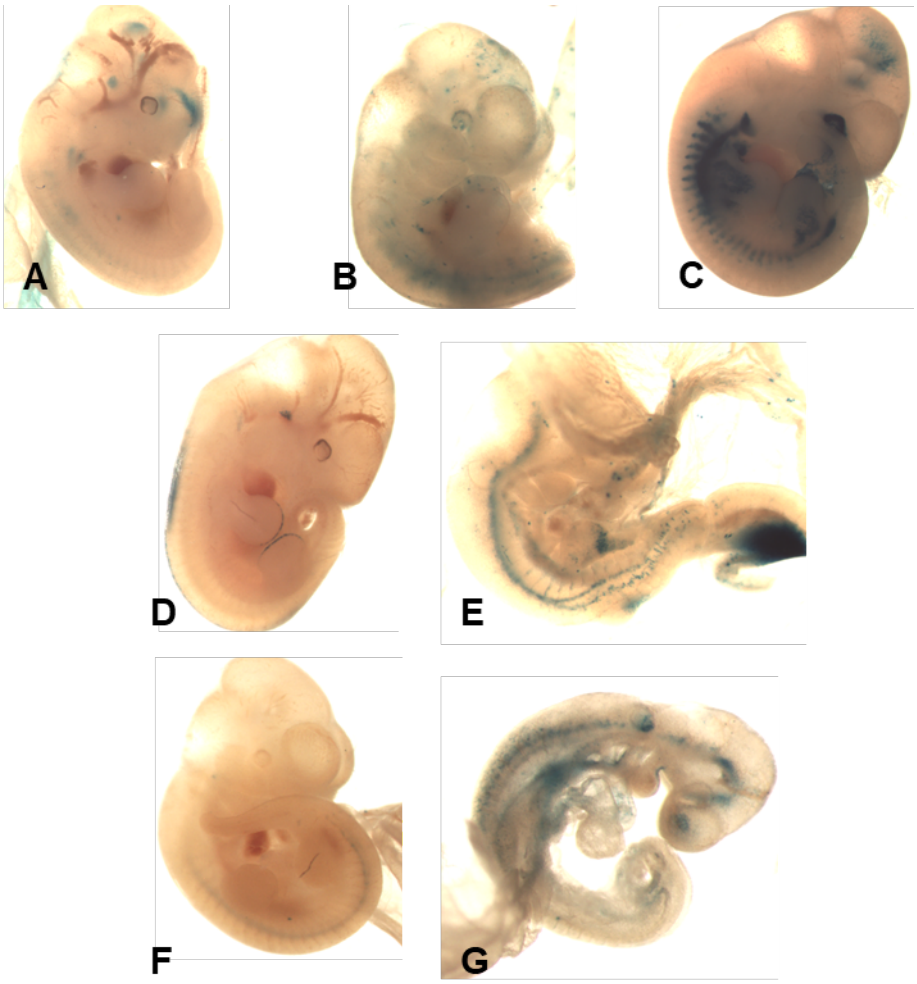


Figure 3.19: X-gal staining of E11.5 F_0 -148kbHsp69LacZ transgenic mice. X-gal staining occurred in several tissues across several founders. Staining of caudal elements occurred in all embryos. There was a developmental delay in embryos B, E and G, which more closely resemble E9.5 embryos.

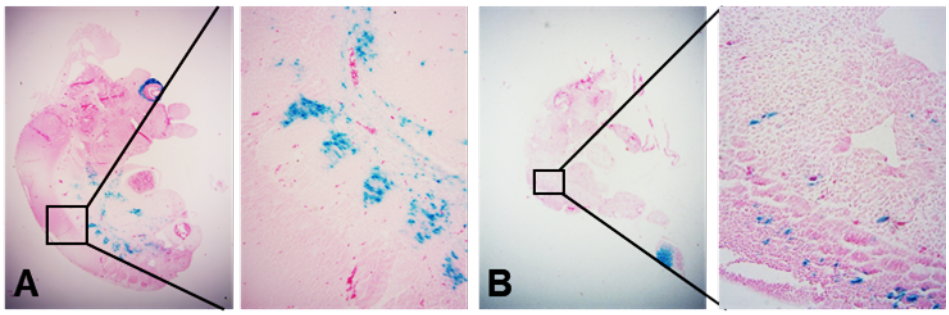


Figure 3.20: Histological sectioning of positively stained founders C and E E11.5 embryos harbouring the -148kbHsp68LacZ construct. Staining of the somites of embryo C was confirmed upon histological sectioning, with strong staining of the tissue throughout the sagittal plane of the embryo (A). The sectioning of Embryo E (B) was punctate across the embryo, the developmental delay made it difficult to compare the tissues in which transgene activity was observed.

3.3.2 -148kb functions at E15.5 in founders

Founder embryos were also used to assay the ability of the -148kb enhancer to drive expression at E15.5. This work was carried out by Dr Ian Li. Three embryos exhibited positive X-gal staining and were imaged (Figure 3.21), with localisation of transgene activity recorded in Table 3.7

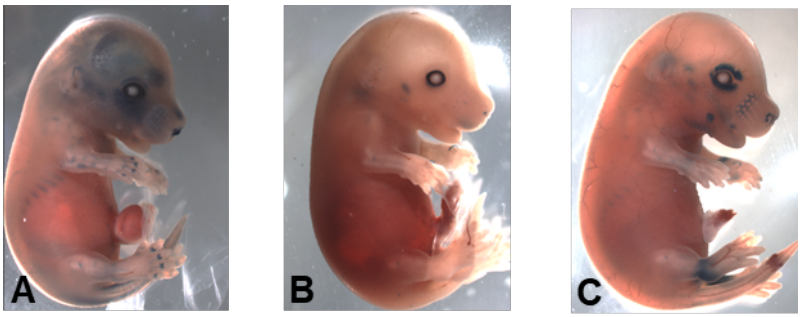


Figure 3.21: Whole mount imaging of founder E15.5 harbouring the -148kbHsp68LacZ construct. Each positive embryo had a different pattern of staining. The ribs and maxillary region stained in all positive F_0. Staining within the limb also occurred within all three founders albeit with differential intensity.

Founder	Whole mount staining intensity	Staining localisation	Intensity of local staining
A	+++	Cranium, ribs, limbs	++++
B	+	Maxillary, ribs, limbs	+++
C	++	Eye, mouth, ribs, wrist, ankle	++++

Table 3.7: X-gal staining and localisation in -148kbHsp68LacZ transgenic founder E15.5. Staining intensity (+ is negligible, ++ diffuse, +++ moderate, ++++strong, +++++intense). Skeletal elements were commonly stained between each founder.

Activity within chondrocytes was observed upon sectioning of all three embryos, albeit in slightly different sub-populations of chondrocytes. There was staining of pre-hypertrophic and hypertrophic chondrocytes in all three embryos, with this being most clearly observed in founder A. There was staining of the interzone of articular joints in Founders B and C such as the elbow of founder C and future joint between basioccipital and vertebral atlas and elbow (Figures 3.22 C and E respectively). Further mesenchymal transgene activity was observed in all three embryos such as surrounding the eye of founder C (Figure 3.22 F). Tissue of the dorsal root ganglion was also positive for transgene activity in Embryo B (Figures 3.22 C and D).

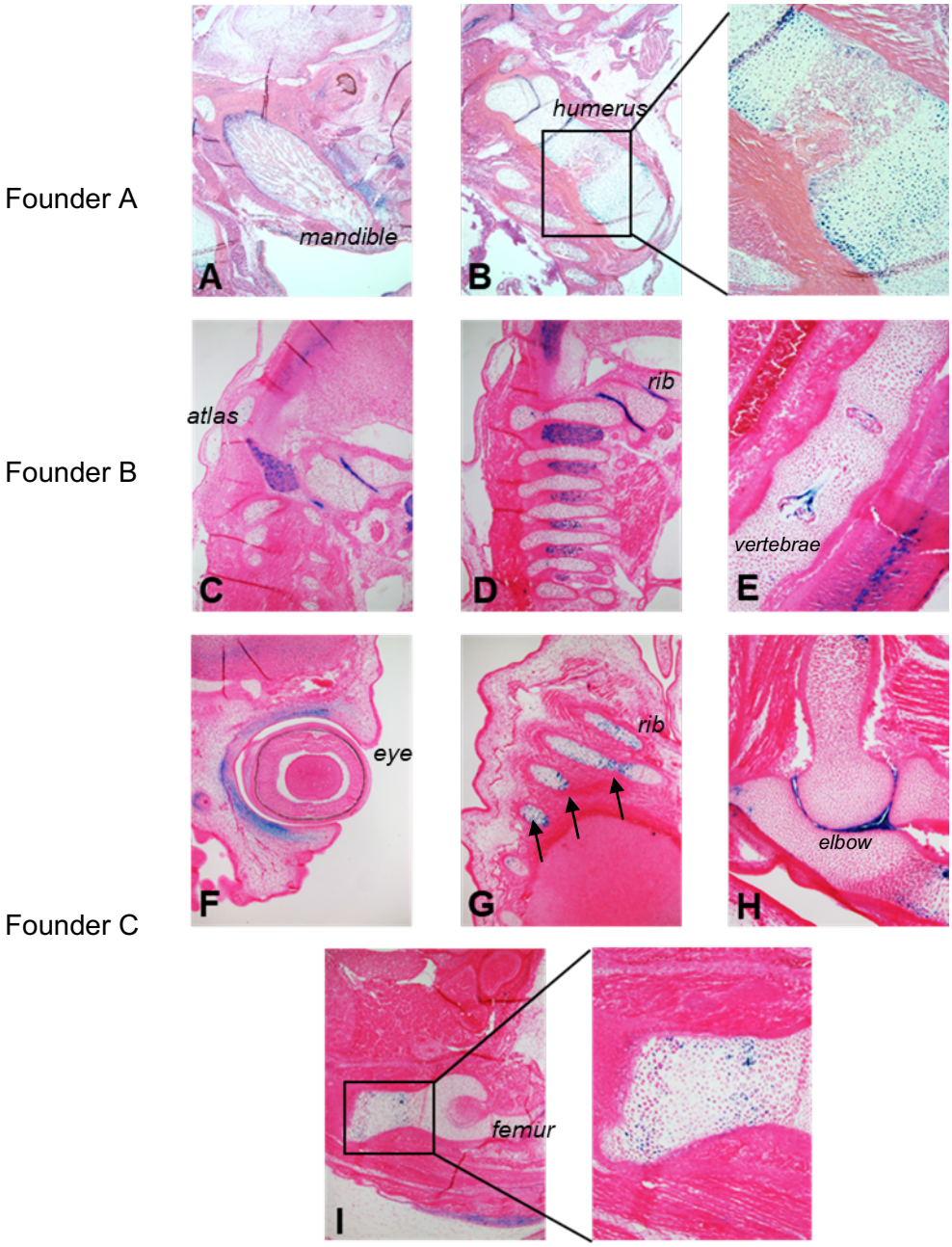


Figure 3.22: Representative histological sections of -148kbHsp68LacZ transgenic founder E15.5. Staining of Embryo A (A and B) occurred most prominently within pre-hypertrophic and hypertrophic chondrocytes, such as within the humerus (B). Perichondrium/periosteum also exhibited modest transgene activity in Embryo A, such as within the mandible (A). There was punctate chondrocyte staining within Embryo B (C-E) within structures such as the vertebrae (C), ribs (D) and tail (E). Pre-hypertrophic chondrocytes were also positively stained within Embryo C (F-I), including within the ribs (arrows, G), humerus (H) and femur (I). Further transgene activity was most notable within the mesenchyme surrounding the eye (F).

The clearest demonstration of hypertrophic chondrocyte expression was observed within the limbs of founder A (Figure 3.22 B). There was an absence of transgene activity within the chondrocytes of the resting and proliferative zones, but strong

staining of pre-hypertrophic zones of the growth plate such as in the humerus (Figure 3.22 B). Osteoblasts within the primary ossification centre were not X-gal stained. Weak staining of the perichondrium and periosteum was also observed such as within the mandible and ribs (Figure 3.22 A and B).

3.4 LacZ expression occurred at several time points in a stable -148kb transgenic mouse line

After preliminary assessment of activity within founders, a stable transgenic line (CTGF148) was propagated from a male founder (F_0 4.3), which was crossed with a wild-type B6CBAF1/J female. Because of the creation of this line, enhancer function was tested at more time-points than for the other enhancers. The expression of the *LacZ* reporter gene was observed using X-gal staining of tissues at several time points during embryonic development and beyond.

3.4.1 -148kb functions at E11.5 in CTGF148

Transgene expression within the CTGF148 line was firstly assayed at E11.5, as illustrated in whole mount images in Figure 3.23. After only two hours (2hr) of exposure to the X-gal substrate, staining was observed in a punctate and superficial manner across embryos, this pattern was much richer after overnight staining with deep blue colour in embryos (Figure 3.23)

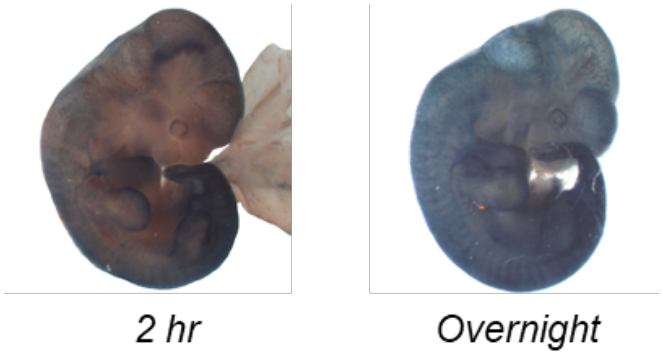


Figure 3.23: Representative whole mount imaging of an X-gal stained E11.5 embryo from the CTGF148 line. After two hours, staining occurred in a punctate, superficial manner globally, tallying with vascular localisation. Further staining was also visible in the heart. This staining pattern intensified after overnight incubation.

Histological sectioning of these embryos reinforced the observation from whole mount imaging that positive X-gal staining most prominently occurred in the vasculature. Throughout each embryo, erythrocytes were observed within

branched blue stained blood vessels. This was well demonstrated in the superficial superior cranium (Figure 3.24 A). Potent transgene activity was observed throughout the sagittal plane of the heart, for example within the branchial arch and primordial tissues of both ventricles and atria (Figure 3.24 B).

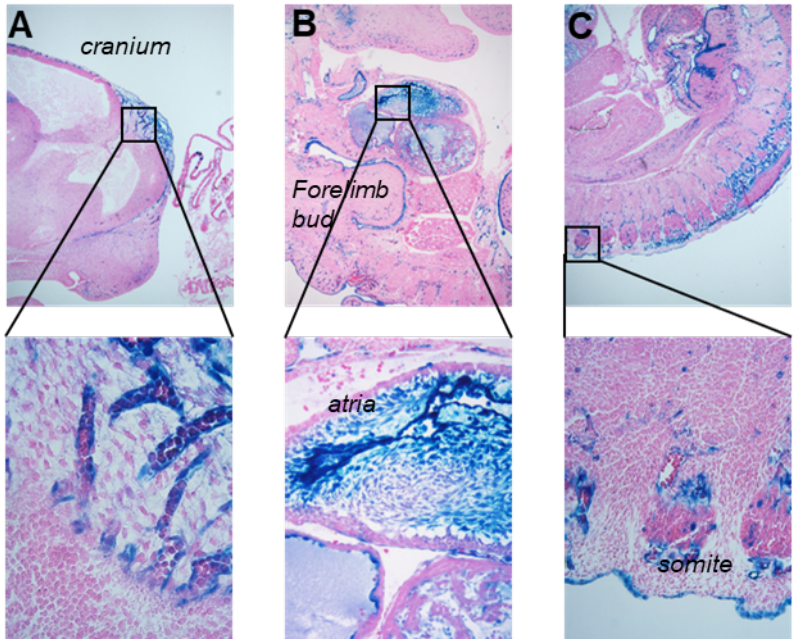


Figure 3.24: Representative histological sectioning of CTGF148 E11.5 embryos. Vascular localisation of staining was prominent, with erythrocytes visible within stained structures. This occurred within the superficial surface of the cranium (A) and trunk (C). Strong blue staining was observed within the heart, particularly within the branchial arch and lining of the atria (B).

Further transgene activity was also observed within the distal edges of the limb buds (Figure 3.24 B). However, there was no staining within the primitive mesenchymal tissue within the limb bud Figure 3.24 B). The somites were negative for staining, but vasculature surrounding them was positive (Figure 3.24C).

3.4.2 -148kb is active at E12.5 in CTGF148

Transgene activity within the CTGF148 line was assessed as E12.5. After overnight X-gal staining, strong positive staining was observed across embryos in whole mount embryos, as demonstrated in Figure 3.25. Staining predominantly occurred in a superficial manner, with blue staining co-localising with blood vessel structures. Rich transgene activity was also observed in the limbs and tail, with dark blue staining of the distal surfaces of the primitive fore and hind paws.

Figure 3.25: Representative whole mount imaging of CTGF148 transgenic mice at E12.5 after overnight X-gal staining. Strong β-galactosidase activity was observed globally, mostly in a superficial manner, for example, staining was not clearly observed within the primitive skeletal anlage.

The pattern of whole mount staining was similar to that observed at E11.5. This was also true upon histological sectioning. Blue staining was most prominent within the superficial vasculature, with rich blue pigmentation of blood vessel structures encasing erythrocytes such as in the tail (Fig. 3.26 B). Staining was also observed within the heart; however this was weaker compared to staining at E11.5 (Figure 3.26), within the endocardial cushions and cells lining the ventricles exhibiting positive staining. (Fig. 3.26 B and C).

Further staining was also observed within primitive cartilaginous tissue. Positively staining chondrocytes were most abundant in the parachordal cartilage of the chondrocranium (Fig. 3.26 A), in addition to the proximal forelimb (Fig. 3.26 D). Positively stained cells within these structures were distributed in a mosaic-like manner. This diffuse manner of staining was also observed in the primitive vertebral column, albeit to a much lesser extent.

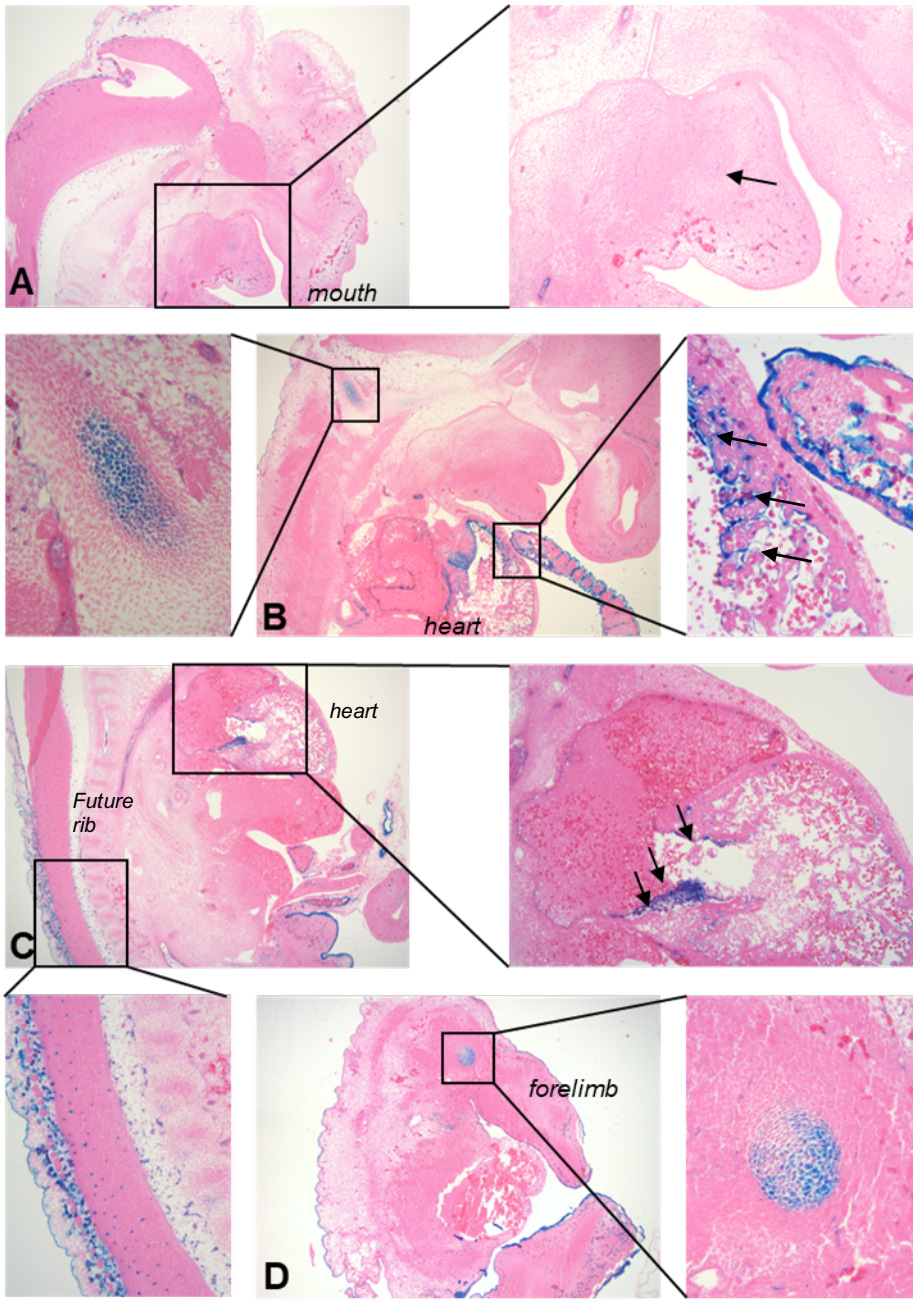


Figure 3.26: Histological sectioning representative of X-gal stained E12.5 from the CTGF148 line. Staining occurred most prominently within superficial vascular structures including within the tail (B) and dorsal dermis (C). Further vascular-related transgene activity occurred within the heart (arrows, B and C). Blue positive cells were also observed within the primitive cartilaginous condensations of the chondrocranium (arrow, A) and scapula (D).

3.4.3 -148kb functions at E13.5 in CTGF148

Transgene expression within the CTFG148 line was also assayed at E13.5. Whole mount imaging was conducted after two hours of X-gal staining, in addition to overnight staining (as illustrated in Figure 3.27). The strongest β-galactosidase activity was detected in the cartilaginous anlage of the skeleton, occurring within the limbs after only two hours of staining.

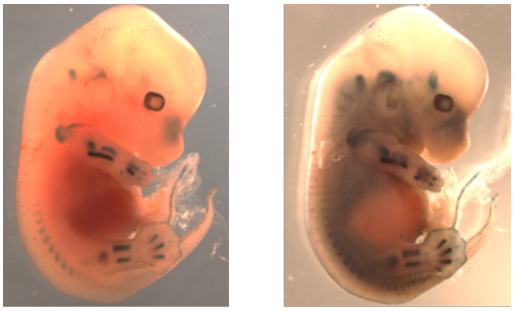


Figure 3.27: Whole mount images of a typical X-gal stained CTGF148 E13.5 embryo at varying staining durations. After two hours of staining, prominent transgene activity was observed within the limbs, caudal vertebrae and nasal bone primordia tallying with the localisation of endochondral cartilage anlage.

After overnight staining incubation, staining was observed in regions corresponding to cartilaginous tissue across each embryo. Additional albeit weak staining was observed in a more superficial manner at the distal edges of the primitive fore and hind paws. This contrasts the potent superficial staining that was seen at E11.5 (Figure 3.23). These observations were reaffirmed upon histological sectioning of the embryos (Figure 3.28).

Chondrocytes exhibited strong β-galactosidase activity in every cartilaginous element, such as within the Meckel's cartilage (Fig. 3.28 A), ribs and vertebrae (Fig. 3.26 B) and future long bones including the tibia and fibula (Fig. 3.28 D).

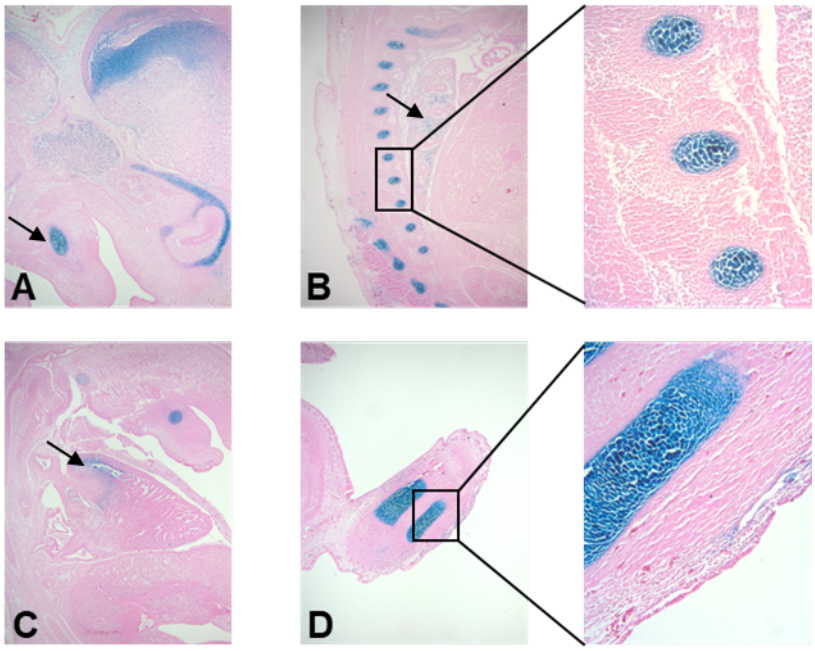


Figure 3.28: Representative histological sections of X-gal stained CTGF148 E13.5 embryos. The most significant positive X-gal staining of cells occurred observed within cartilaginous tissues, including the Meckel's cartilage (arrow, A) and chondrocranium (A), vertebrae and ribs (B) and limb such as in the primitive tibia and fibula (D). Faint staining was also observed within the atria of the heart (C).

The strong X-gal staining observed within vascular elements at E11.5 and E12.5 (Figures 3.24 and 3.26 respectively) was not repeated at E13.5. This was most notably observed in the heart (Figure 3.26 C), where staining was present in the atria. Superficial staining of the vasculature as seen at E11.5 was much decreased at E13.5. Additional staining was also observed within the brain (Figure 3.28 A) and lung (arrow, Figure 3.28 B).

3.4.4 -148kb functions at E15.5 in CTGF148

The capacity of -148kb to drive the expression of *lacZ* in CTGF148 was examined at E15.5. Strong expression of the transgene was observed in a cartilage specific manner, as demonstrated in Figures 3.27 and 3.28. Potent β-galactosidase activity was demonstrated after a short duration of X-gal staining (2 hr, Figure 3.29). Upon this initial stage of imaging, strong blue staining was observed within all limbs and each primordial skeletal element within. Further staining was also observed within the chondrocranium and ribs.

Figure 3.29: Whole mount imaging of X-gal stained E15.5 embryos from the CTGF148 line at varying stain durations. Strong blue staining was observed within the limbs and tail after two hours of staining, with weaker staining observed within the chondrocranium and ribs. After overnight staining, cartilaginous tissue has stained intensely on a global scale. This was highlighted by soft tissue clearing of overnight stained embryos, which demonstrated intense staining of every cartilaginous structure.

Embryos were re-imaged after overnight staining incubation (overnight, Figure 3.29). The staining pattern occurred in the same locations as at two hours, however staining was much more intense than after two hours. Embryos were then cleared to remove soft tissue before further imaging (cleared, Figure 3.29). Every cartilaginous tissue within the embryos exhibited potent β-galactosidase activity. Osseous tissues that arise through intramembranous ossification such as the frontal and parietal bones of the cranium were negative for staining.

Histological sectioning of embryos further reinforced the observation of intense cartilaginous staining, as illustrated in Figure 3.30. Moreover, staining was confined to chondrocyte cells within positive tissues. Soft tissue structures such as the eye (Figure 3.30 D) did not contain positively stained blue cells. There was no staining of cardiac tissue as seen at E11.5, E12.5 and E13.5, as shown in the heart (Figure 3.30 H) in which there were no transgene positive cells.

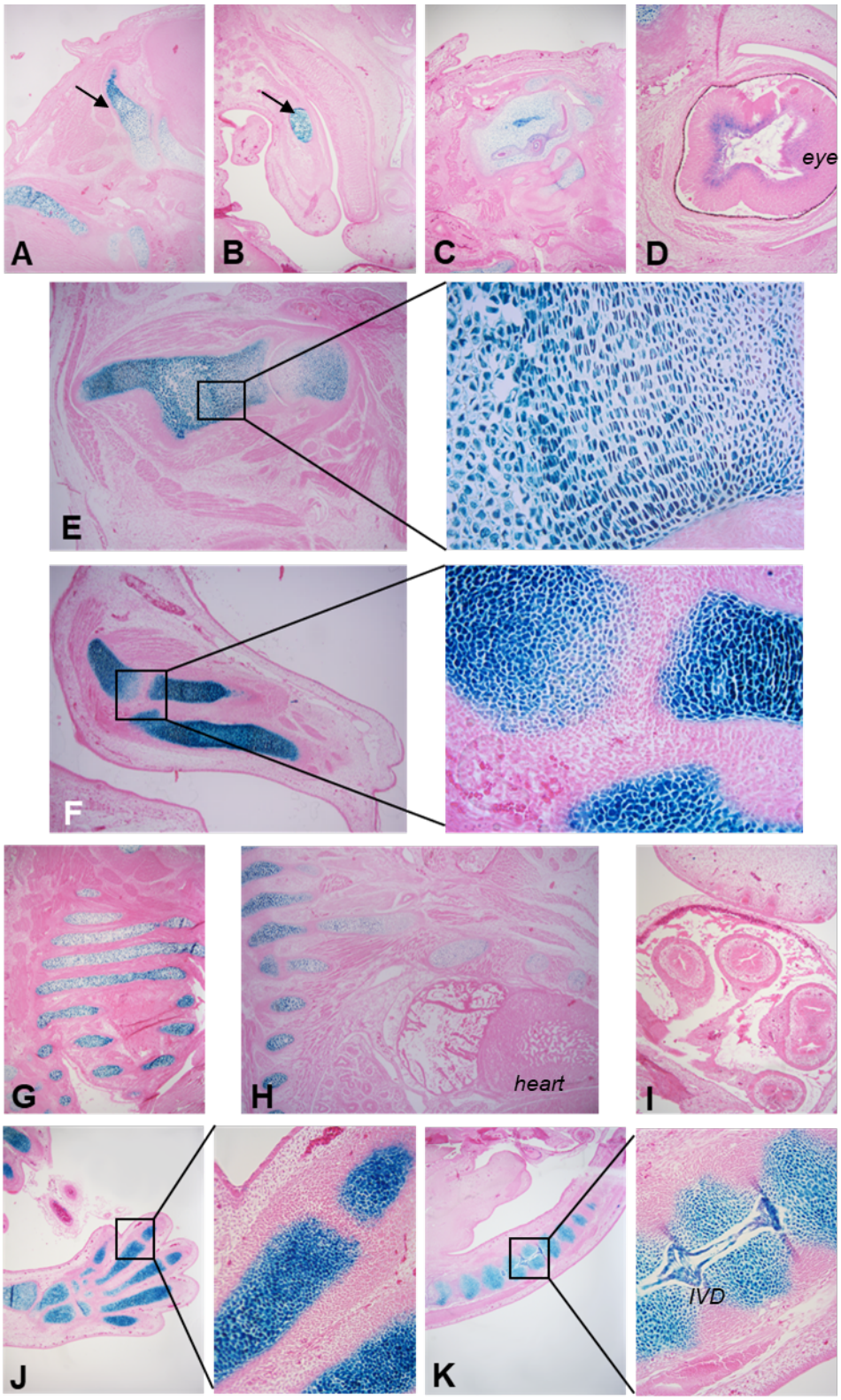


Figure 3.30: Representative histological sections from CTGF148 E15.5 embryos. Staining occurred in chondrocyte cells throughout the sagittal plane of embryos. Hyaline cartilage anlage of the exoccipital bone (arrow, A), parietal (C), scapula (E), humerus, radius and ulna (F), ribs (G and H), hind paw (J) and tail (K) bones all exhibited strong transgene activity. Further cartilaginous β-galactosidase activity was observed within the Meckel's cartilage (B) and intervertebral disc (K). Soft tissue elements such as the eye (D), heart (H) and mid-gut loop (I) were all negative for staining.

Sectioning revealed that not every chondrocyte demonstrated intense staining. For example, the frontal portion of the ribs (Fig. 3.30H) contained β-galactosidase positive cells, but this was not homogenous across each rib structure, with punctate staining pattern of cells within some ribs. This was echoed in the shoulder joint (Fig. 3.30E), where more diffuse staining was observed in the resting zone of chondrocytes near to the shoulder joint, with greater proportion of positive cells in more mature chondrocytes, towards the future bone diaphysis of the humerus and centre of the scapula.

3.4.5 -148kb is active at E17.5 in CTGF148

Transgene expression driven by the -148kb enhancer within the CTGF148 line was also examined at E17.5. The most potent β-galactosidase activity occurred within tissue undergoing endochondral ossification, including the scapula (Figure 3.31 B), humerus (Figure 3.31 C), ribs (Figure 3.31 D) sternebrae (Figure 3.31F), femur, tibia and fibula of the hind limb (Figures 3.31 G and H) and phalanx (Figure 3.31I). Unlike the staining of cartilage that occurred at E15.5, there was heterogeneity in the extent of staining observed within endochondral tissue. Moreover, this occurred in line with the differentiation state of chondrocytes; blue X-gal staining intensity was depleted with increasing proximity to the primary ossification centre (ossification centres are marked with arrowheads across Figure 3.29). Further chondrocyte based transgene activity also occurred within the chondrocranium (Figure 3.31 A) and xiphoid (Figure 3.31 E).

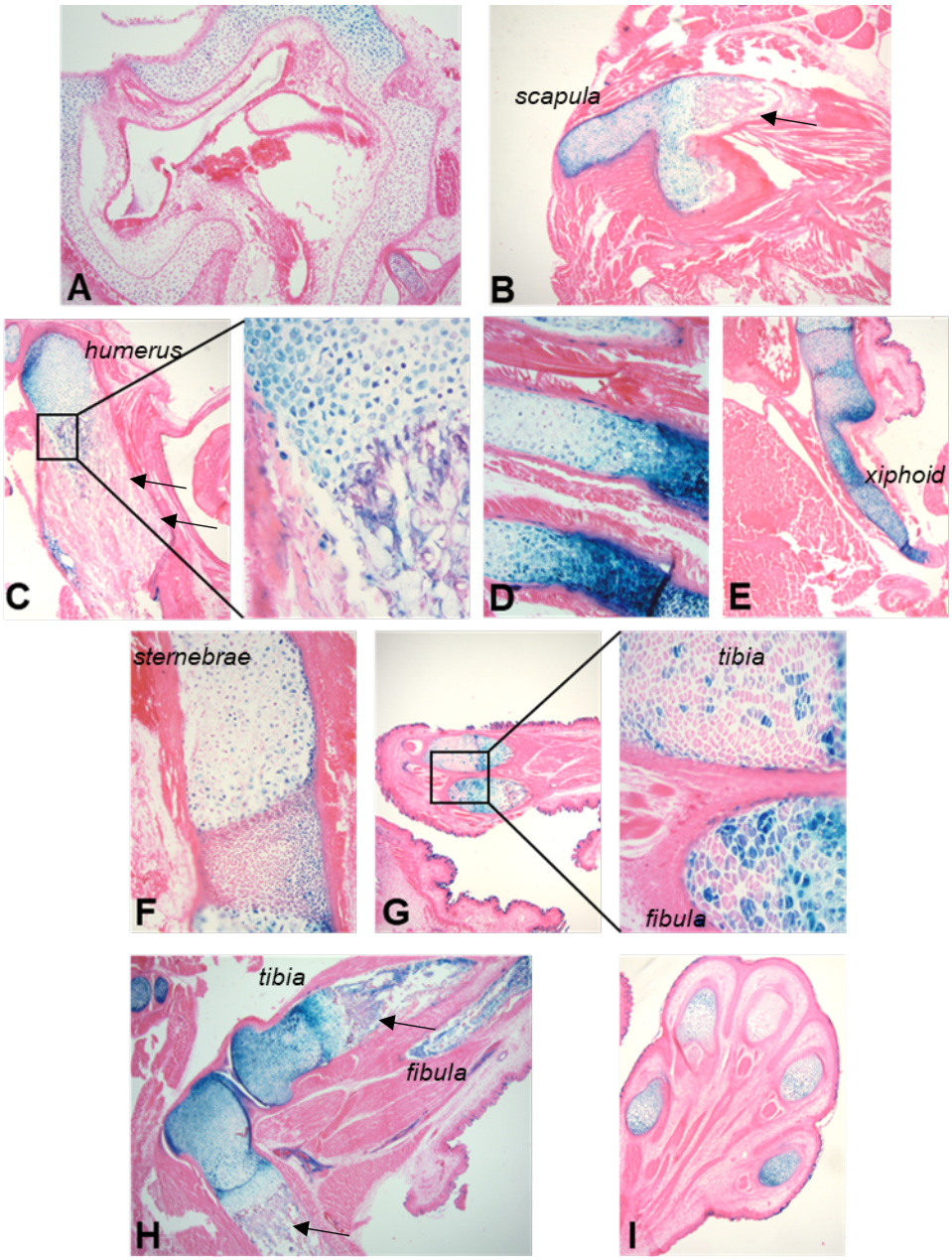


Figure 3.31: Representative histological sections from E17.5 CTGF148 embryos. The most profound transgene activity was observed within the chondrocytes of cartilaginous tissue, such as within the chondrocranium (A), scapula (B), humerus (C), ribs (D), xiphoid (E), sternebrae (F), hind limb (G and H) and hind paw (I). However, the distribution of staining of chondrocytes occurred in a heterogeneous manner across cartilage, with richer transgene activity in distal location from ossification centres. This was most strikingly observed in the humerus (C) and femur and tibia (H). Ossification centres are highlighted with arrows.

3.4.6 CTGF148 mice exhibit β-galactosidase activity in early adulthood

Mice from the CTGF148 line were taken for tissue in early adulthood at nine weeks of age. Staining for β-galactosidase activity revealed strong transgene expression within cartilaginous structures, as demonstrated in whole mount imaging of soft tissue cleared skeletal structures (Figure 3.32).

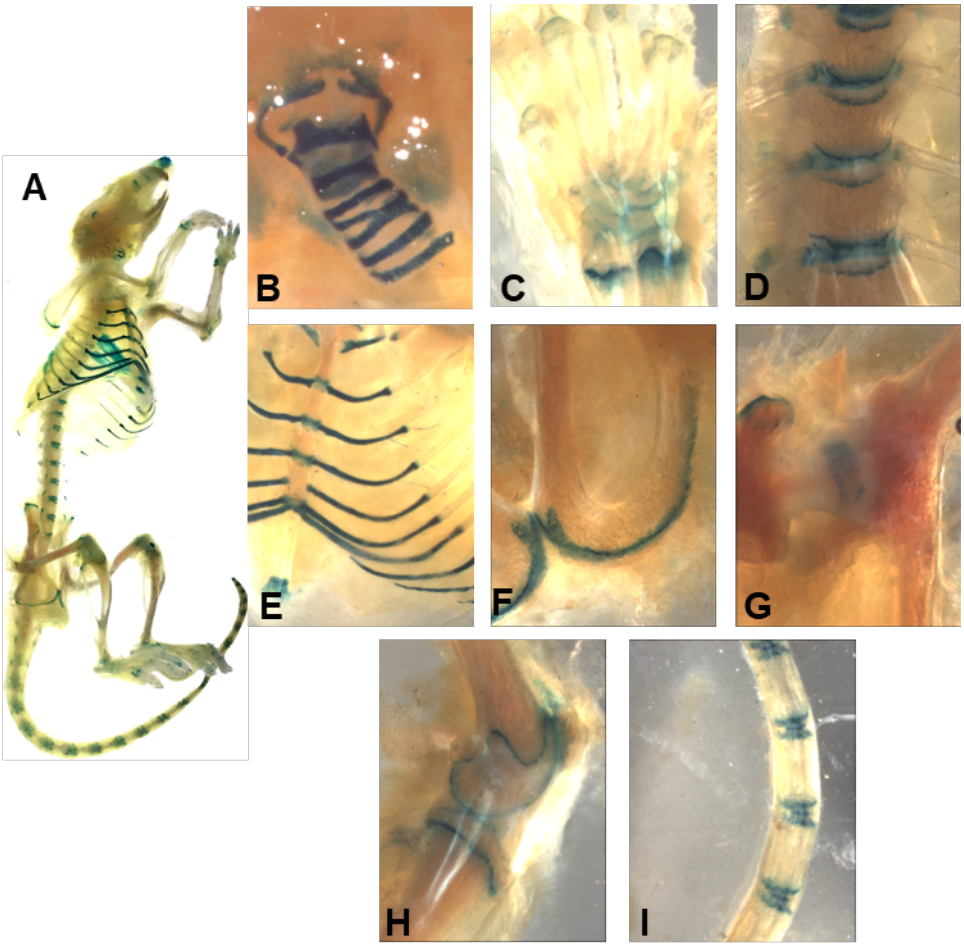


Figure 3.32: Representative whole mount imaging of soft tissue cleared, X-gal stained nine week old CTGF148 mice. Cartilage structures demonstrated strong global staining (A). This included the trachea (B), costal cartilage (E) in addition to the articular surfaces of the hip (G) and knee (H) joints. Growth plate cartilage also exhibited strong staining within the wrist, thoracic vertebrae (D) femur and tibia (H) and caudal vertebrae (I). The fibrocartilage of the intervertebral disc also exhibited transgene expression.

The chondrocyte-specific nature of β-galactosidase activity at this time-point was reinforced upon histological sectioning, which revealed staining predominantly within hyaline cartilage structures such as the costal cartilage (Figure 3.33 A), articular cartilage of the vertebral endplates (Figure 3.33 B), femoral head (Figure 3.33 C) knee (Figure 3.33 D); in addition to the epiphyseal growth plate such as within the tibia (Figure 3.33 D). The transgene was also active in fibrocartilage structures including the annulus fibrosus of the intervertebral discs (Figure 3.33 B) and knee meniscus (arrow, Figure 3.33 D). There was an absence of staining within osseous structures on a global scale, with cartilaginous growth plates the only location within bones where positive X-gal staining occurred.

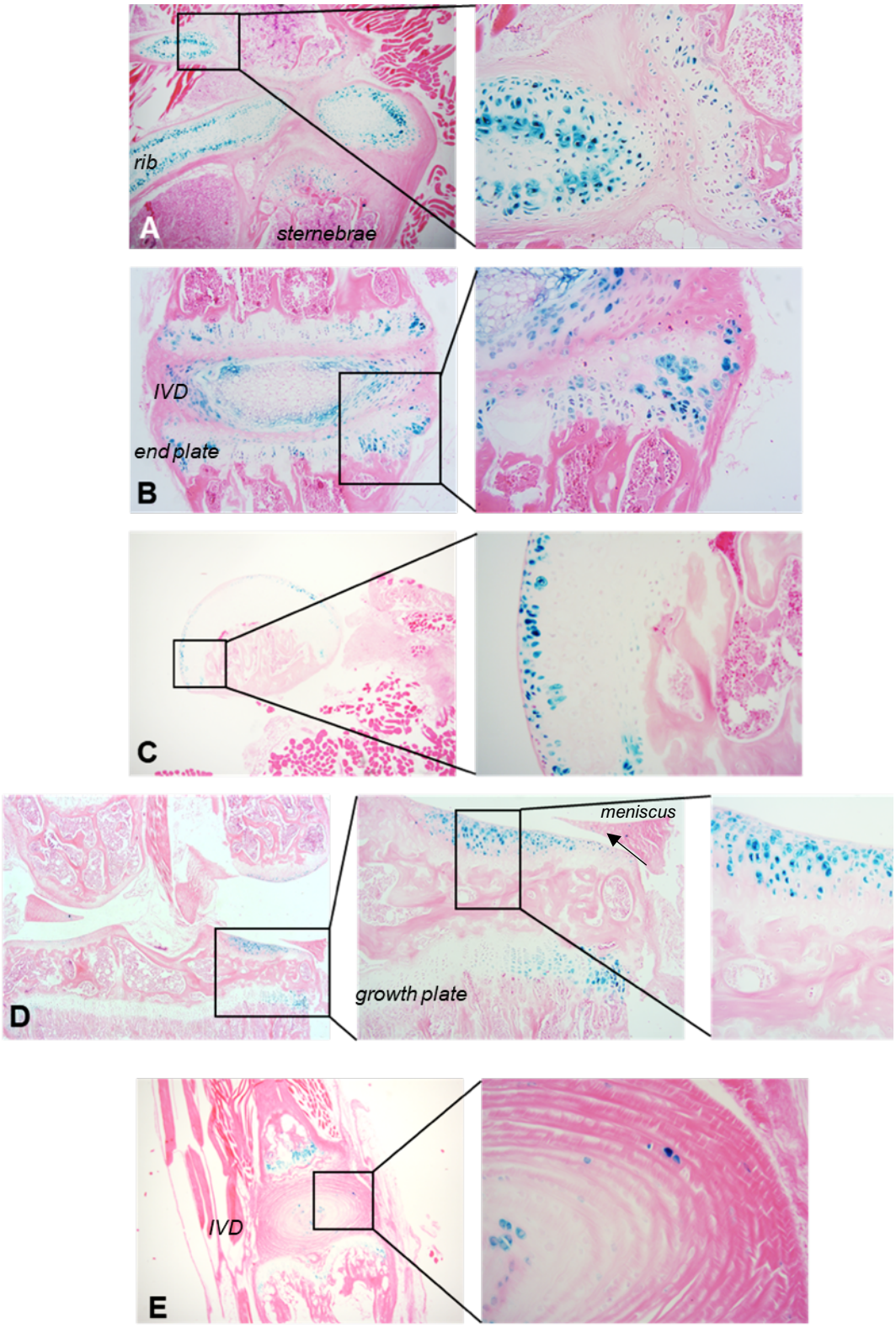


Figure 3.33: Histological sections of X-gal stained CTGF148 mouse at nine weeks of age Transgene activity exclusively occurred within chondrocytes. Hyaline cartilage of the costal cartilage (A), articular cartilage of the femoral head (C) and knee (D) and epiphyseal plates globally all exhibited positive staining. Further chondrocyte based transgene activity was observed within the fibrocartilage of the intervertebral disc such as within the annulus fibrosus of the thoracic vertebrae (B) and tail (E).

3.4.7 -148kb drives transgene activity during mature adulthood in CTGF148 mice

Transgene function within the CTFG148 line was also tested in mature adulthood at approximately five months of age. Whole mount imaging of X-gal stained tissue revealed cartilage specific staining (Figure 3.34).

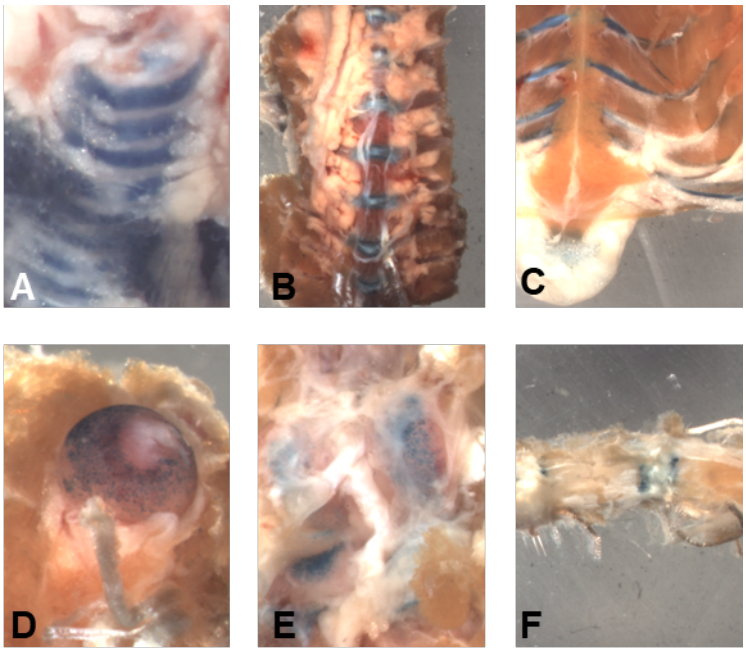


Figure 3.34: Whole mount imaging of X-gal stained CTGF148 mice at five months of age. Staining was observed within many structures composed of hyaline cartilage, such as the trachea (A), costal cartilage (C) and femoral head and knee articular cartilage (D and E respectively). Further cartilaginous transgene activity was observed within the intervertebral disc (B and F).

Strong staining of hyaline cartilage occurred throughout the animal, within structures including the trachea (Figure 3.34 A and Figure 3.3 5A), costal cartilage (Figure 3.34 C) articular joint surfaces of the femoral head (Figure 3.34 D and Figure 3.35 B) knee (Figure 3.34 E and Figure 3.35 D and E), in addition to the vertebral endplate of the thoracic and tail vertebrae (Figures 3.34 B and F, Figure 3.35 C). The staining observed both in whole mount and histological sectioning was weaker at this time-point than in early adulthood (Figure 3.32). For example, staining of the femoral head was more punctate in mature adulthood, with few blue articular chondrocytes in histological sections (Figure 3.35 B).

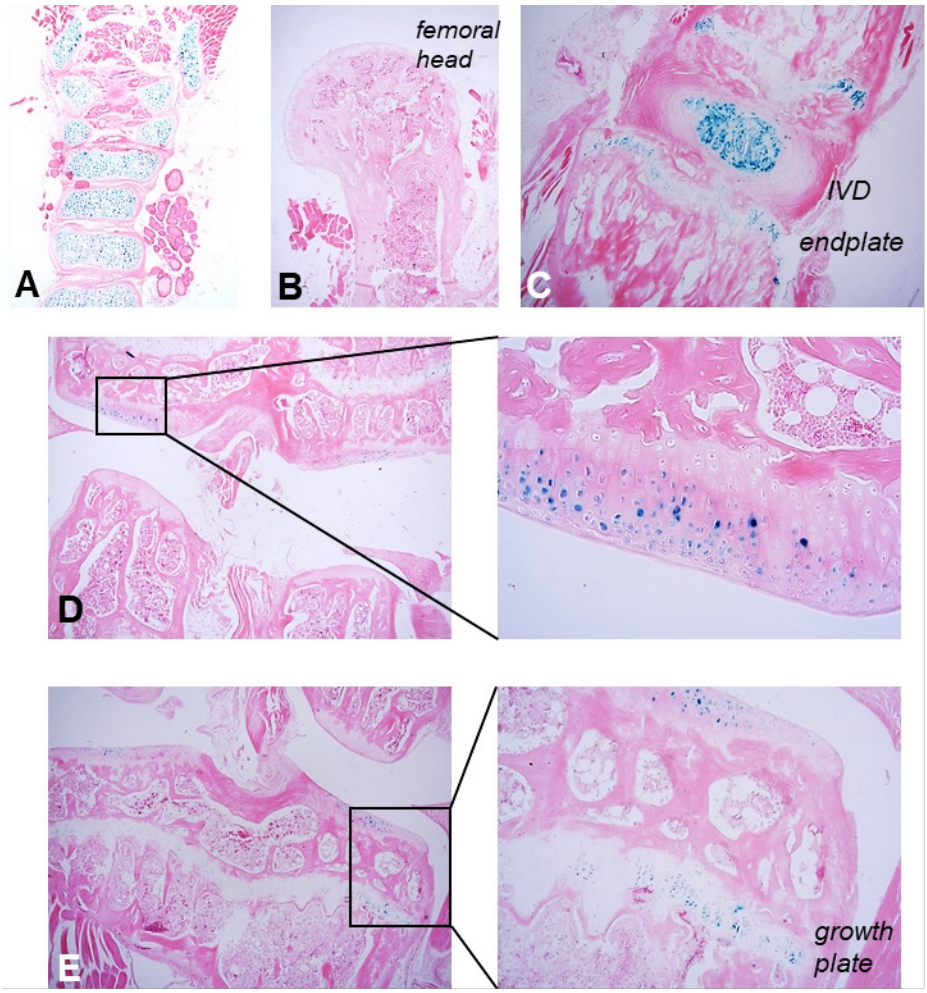


Figure 3.35: Histological sectioning of X-gal stained tissue from five month old CTGF148 mouse. Staining was observed in a chondrocyte specific manner. Punctate distribution of staining was observed within chondrocytes of hyaline cartilage within the trachea (A), in addition to growth plate and articular surfaces of the femoral head (B), femur (D) and tibia (E). The fibrocartilage structure of the nucleolus pulposus of the intervertebral disc (C) also exhibited transgene activity.

3.5 A truncated version of -148kb functions in embryonic development and adulthood

A reductionist approach was used in order examine the sequence required for -148kb enhancer function. Upon visualisation of the enhancer in the UCSC Genome Browser (Figure 3.36), peaks for enhancer associated chromatin attributes (Figure 3.36 C) and conservation of the sequence (Figure 3.36 D) are split into two sub-regions. A ScaI restriction enzyme site between these two peak areas was used to cut the -148kbHsp68LacZ construct (red dashed line, Figure 3.36). This created the -148kb_short construct (Figure 3.36 A) which encompassed the 3' 1529bp sequence of the enhancer. Transgenic mice were created as outlined previously, with the resultant line being named **-148kb_short.**

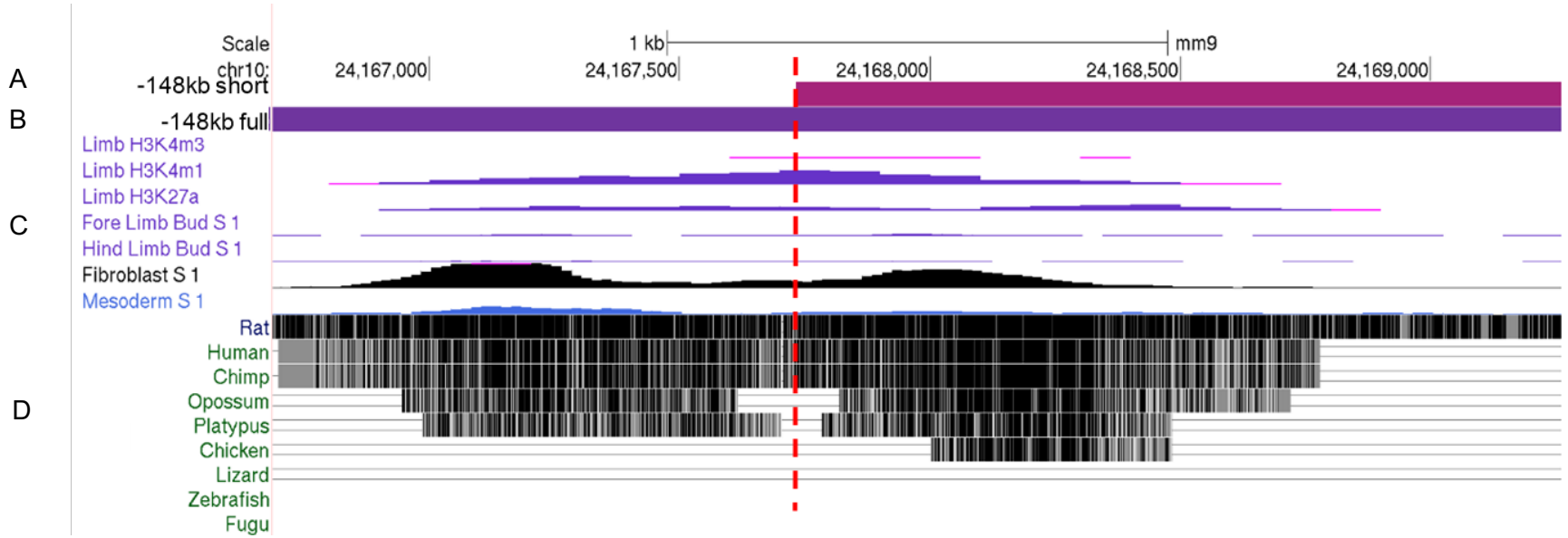


Figure 3.36: Visualisation of the sequences used to create the -148kb_short transgenic mouse line. The full -148kb sequence (B) of the -148kbHsp68LacZ construct was cut (red dashed line) on the basis of a split between peaks for enhancer associated chromatin attributes (C) and sequence conservation (D). This created the -148kb_short construct used to create the -148kb_short line.

Mice were genotyped upon weaning with a total of five founders positive for the transgenic construct. Staining for β-galactosidase activity was conducted on F_0 and F_1 mice originating from each founder. For the generation of F_1, founders were crossed with wild-type littermates. Line variants originated from each founder A-E, and are denoted accordingly; i.e. -148kb_short A denotes tissue from founder A, or its progeny.

3.5.1 -148kb_short drives transgene activity at E15.5

The capacity of the -148kb_short construct to drive the expression of *LacZ* was first examined at E15.5 for each of the F_0 variants of the transgenic mice. Whole mount examination of X-gal stained embryos was carried out, with results recorded in Table 3.8, with representative images in Figure 3.37.

Founder	Whole mount staining intensity	Staining localisation	Intensity of local staining
A	++++++ and +++	A_1; global staining A_2: cranium	++++++
B	+++	cranium	+++
C	++++	Cranium , ribs, limbs	+++++
D	++++++	Global	++++++
E	-	n/a	-

Table 3.8: Whole mount image staining intensity of -148kb_short E15.5 that were positive for X-gal staining. Staining intensity (+ is negligible, ++ diffuse, +++ moderate, ++++strong, +++++intense). The embryos in two transgenic line variants exhibited intense global staining.

Embryos from founders A and D (Figures 3.37 A and D respectively) exhibited intense blue X-gal staining across the whole of each positive embryo, aside for one embryo stemming from -148kb_short A, in which intense staining only occurred in the cranium (Figure 3.37 A_2). The main difference between embryos from these two founders in whole mount imaging was that the eye was positively stained for -148kb_short D, but not A. Positive X-gal staining for E15.5 descendants from founders B and C was more constrained. For example, staining occurred in the cranium for both, yet in E15.5 of founder B staining was diffuse throughout the cranium (Figure 3.37 B), whereas for -148kb_short C, staining overlapped with the location of the frontal and parietal bones (Figure 3.37 C).

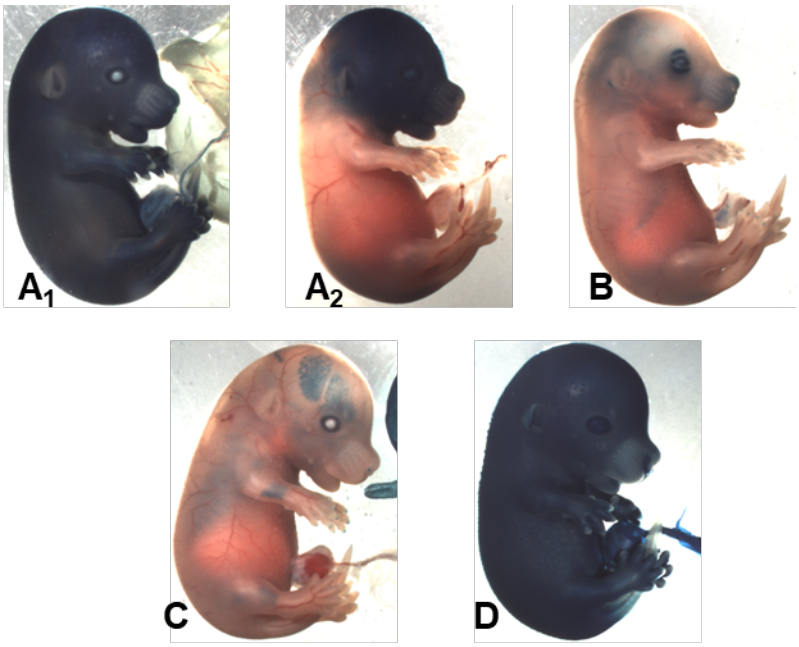


Figure 3.37: Whole mount images of X-gal stained E15.5 from stemming from four founder -148kb_short transgenic mice. Two patterns were observed in the embryos from founder A (A_1 and A_2). E15.5 stemming from -148kb_short A (A_1) and D (D) demonstrated very intense global staining. E15.5 from founder B demonstrated staining within the cranium and midsection. Embryos from founder C exhibited staining that tallied with the parietal bone in the skull, ribs and diaphysis of the long bones of the limbs.

No transgene activity was observed whole mount in embryos derived from founder E. Hence, no further examination of transgene activity was carried out for this founder. Histological sectioning of tissues was carried out for -148kb short_ A to D, enabling greater understanding of the tissue and cellular localisation of transgene activity.

3.5.1.1 Histological examination of β-galactosidase activity in -148kb_short A

The intense blue staining of E15.5 -148kb_short A in whole mount was due to staining within sub-dermal mesenchyme. The epidermis was devoid of staining. However, fibroblastic cells of the reticular layer of the dermis were positive for transgene activity. This was clearly visible within the mouth (Figure 3.38 A) fore paw (Figure 3.38 F) and dorsal region (Figure 3.38 G).

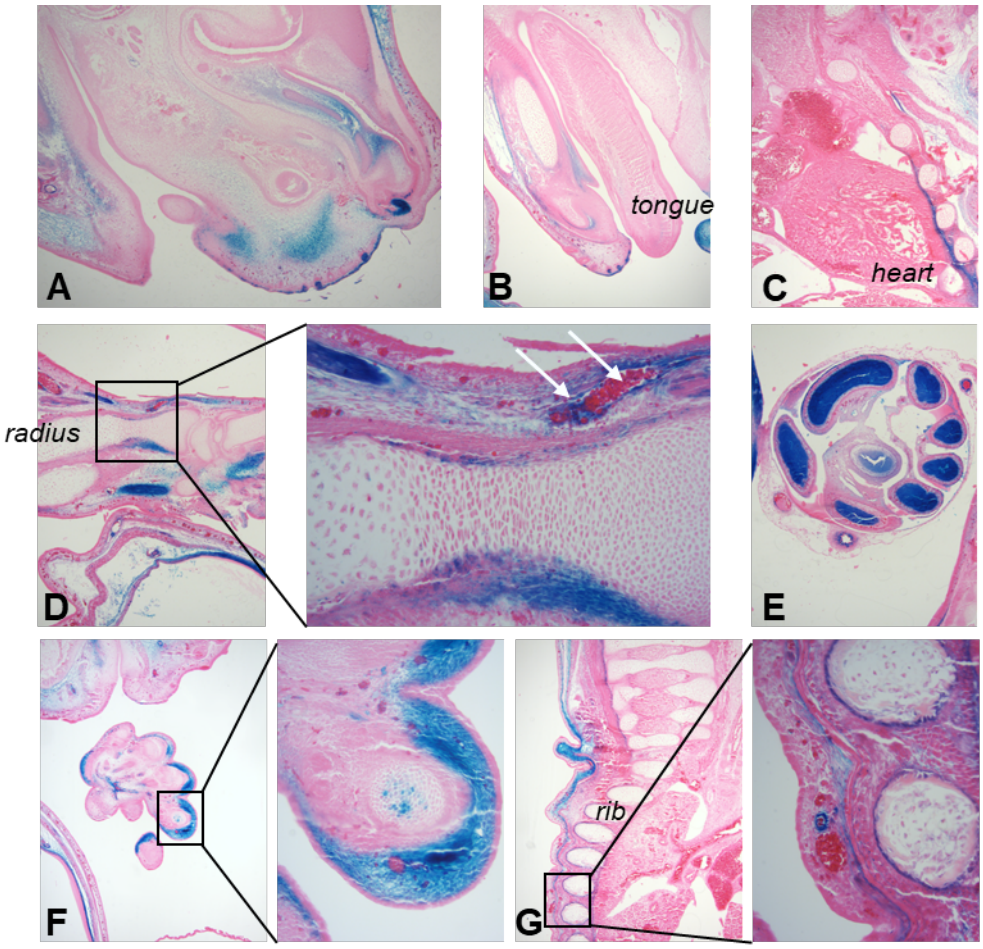


Figure 3.38: Representative histological sections of X-gal stained E15.5 embryos from derived from founder A of the -148kb_short line. Profound subdermal localisation of β-galactosidase activity was observed on a global scale, as demonstrated in the craniofacial region (A), fore paw (F) and dorsal region (G). Primitive skeletal tissues were largely negative for staining, for example no staining was observed within the Meckel's cartilage (B). There was however, some staining of chondrocytes within the phalanges (F) and within the bone collar region of the radius, in addition to discrete staining surrounding the ribs (G). Although there was no staining within the heart (C), there was some staining within blood vessels, which was readily distinguishable due to the presence of erythrocytes (white arrows, D). Further staining occurred in the lumen of the midgut loop (E).

Soft tissue organs such as the heart (Figure 3.38 C) did not demonstrate β-galactosidase activity, although the lumen of the midgut loop was strongly positive for X-gal staining (Figure 3.38 E). Staining was also observed within blood vessels, with blue cells and tissue surrounding erythrocytes (white arrows, Figure 3.38 D). Transgene activity was largely absent within the musculoskeletal system, for example the Meckel's cartilage and surrounding osseous material of the mandible were negative for any X-gal staining. However, there was limited staining the bone collar of the radius (Figure 3.38 D), of chondrocytes within the phalanx of the fore paw (Figure 3.38 F) with some staining and periosteal region of the ribs (Figure 3.38 G).

3.5.1.2 Histological examination of β-galactosidase activity in -148kb_short B

The strong transgene activity within the cranium of -148kb_short B at E15.5 in whole imaging was echoed in histological sectioning, with blue positive stained tissue predominantly occurring within the mouth and nose (Figure 3.39 A). There was no staining within chondro-osseous structures such as the ribs (Figure 3.39 B) or hind limb (Figure 3.39 C).

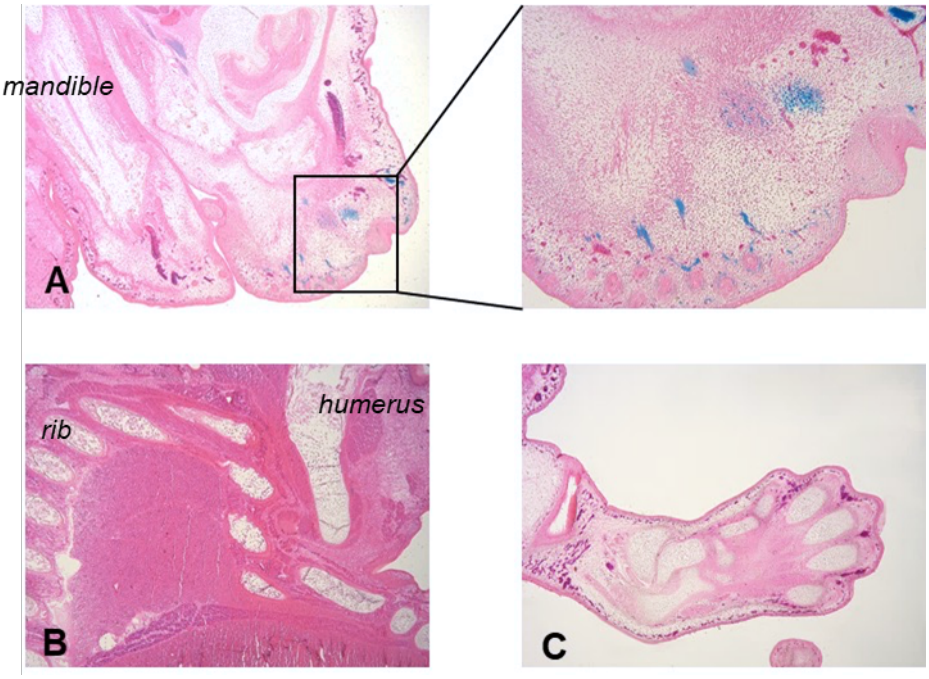


Figure 3.39: Representative histological section images from X-gal stained E15.5 descended from -148kb_short founder B. X-gal staining positive cells were localised predominantly within the soft tissue of the cranium (A). There was an absence of staining within soft tissues such as the heart (B), in addition to skeletal elements such as the ribs, humerus (both B) or hind limb and paw (C).

3.5.1.3 Histological examination of β-galactosidase activity in -148kb_short C

Histological sectioning of E15.5 -148kb_short C revealed highly specific β-galactosidase activity. There was no staining of any soft tissue organ, including the heart (Figure 3.40 E). Staining did not occur in chondrocytes, with cartilaginous structures such as the Meckel's cartilage (Figure 3.40 B) or hind paw (Figure 3.40 F) containing no blue, positively X-gal stained cells. Transgene activity was confined to the periosteal regions of osseous anlage such as within the mandible (Figure 3.40 B), radius (Figure 3.40 C), ribs (Figure 3.40 D) and tibia (Figure 3.40 G).

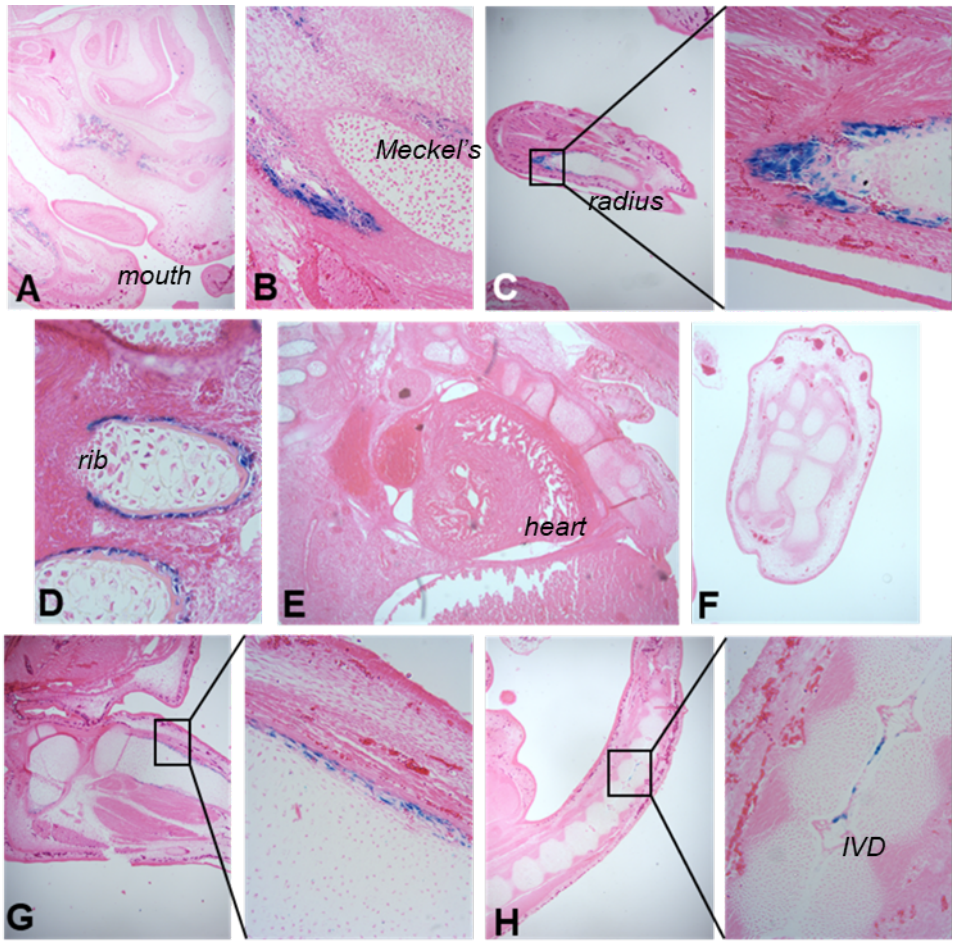


Figure 3.40: Histological analysis of X-gal stained E15.5 progeny of founder C of -148kb_short. Staining was co-localised with primitive periosteal tissue such as within the maxillary (A), mandible (B), fore limb (C), ribs (D) and tibia (G). Soft tissues such as the heart (E) were negative for staining. Further weak staining was observed within the vertebrae of the tail (H).

3.5.1.4 Histological examination of transgene activity in -148kb_short D

Transgene activity occurred in tissue of mesenchymal origin. Positively X-gal stained chondrocytes were present in the cartilage anlage of several bones including the exoccipital and atlas (Figure 3.41 A) and femur and tibia (Figure 3.41 F). The variability of staining intensity within cartilaginous tissues is highlighted in the femur (Figure 3.41 F) with the proportion of blue stained cells changing with proximity to the diaphysis of the future long bone. In addition, other cartilaginous structures such as the temporal bone anlage (Figure 3.41 A) and Meckel's cartilage (Figure 3.41 B) were negative for β-galactosidase activity. Transgene activity was also observed within the mesenchymal tissue of the reticular layer of the dermis, tissue surrounding the follicles of the vibrissae (Figure 3.41 B) and midgut loop (Figure 3.41 G).

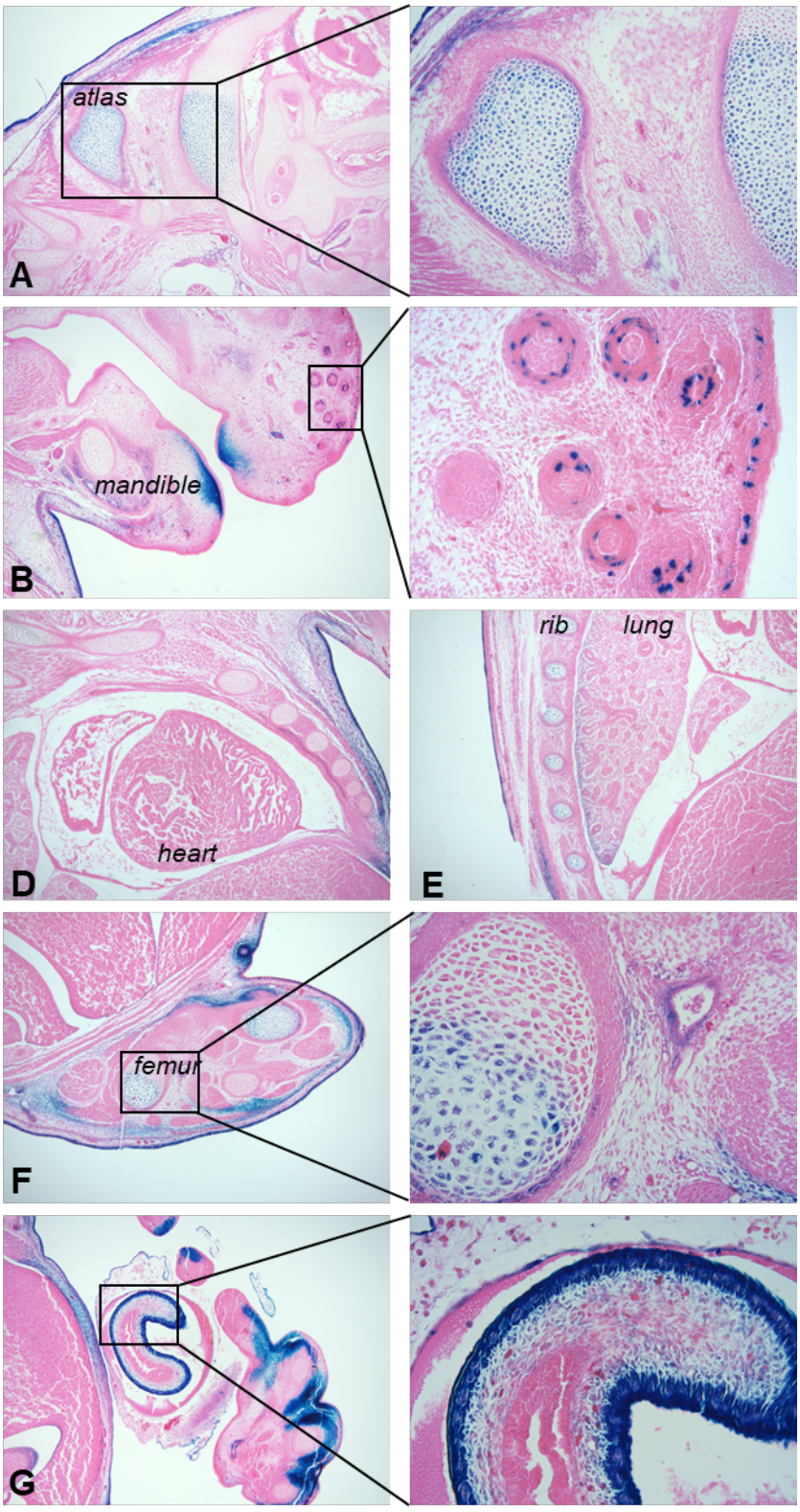


Figure 3.41: Representative images of histological analysis of X-gal stained E15.5 stemming from founder D of -148kb_short. Limited X-gal staining was observed within chondrocytes of future osseous tissues such as in the exoccipital bone and atlas (A), in addition to ribs (E), femur and tibia (F). Other cartilaginous structures including the Meckel's cartilage (B) were negative for transgene activity. Strong staining occurred in the dermis, in addition to mesenchymal tissues, including within the sub-dermal layer such as within the lips (B), the hind limb (F) in addition to the mesenchyme of the primitive vibrissae follicles (B) and midgut herniation (G). The heart (D) and lung (E) were negative for transgene activity.

3.5.2 -148kb_short founder B exhibited negligible transgene expression in adulthood

LacZ transgene activity was assessed in mature adulthood, at approximately five months for -148kb_short founder B. The only location in which transgene activity was observed was the heart, as demonstrated by whole mount and histological sectioning of the heart imaging (Figure 3.42).

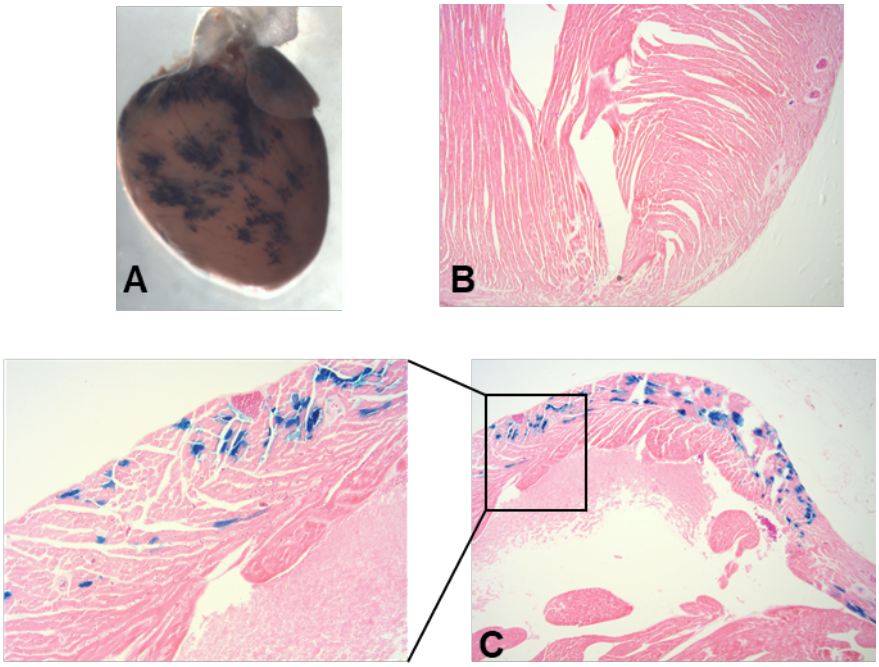


Figure 3.42: Whole mount images of X-gal stained tissue from -148kb_short founder B aged five months. Staining was observed across the surface of the heart in a heterogeneous manner in whole mount imaging (A) and sectioning C).

3.5.3 Potent *LacZ* activity was observed in -148kb_short founder C in adulthood

Staining for β-galactosidase activity was carried out on the tissue of -148kb_short founder C in mature adulthood at approximately six months of age. Intense staining was observed in several tissues, as demonstrated in Figure 3.43. Muscular tissue demonstrated potent transgene activity for example the intercostal (Figures 3.43 D and E) and gastrocnemius (Figure 3.43 G). This was not true of cardiac musculature, with rich X-gal staining of the atria (arrow, Figure 3.43 B), yet superficial staining across the heart was not in a uniform manner (Figure 3.43 C). Further soft tissue based transgene activity was observed in the kidney with punctate superficial staining. There was no clear staining of osseous or

cartilaginous tissue, with no blue staining of the articular surfaces of the hip or knee (Figures 3.43 F and G respectively).

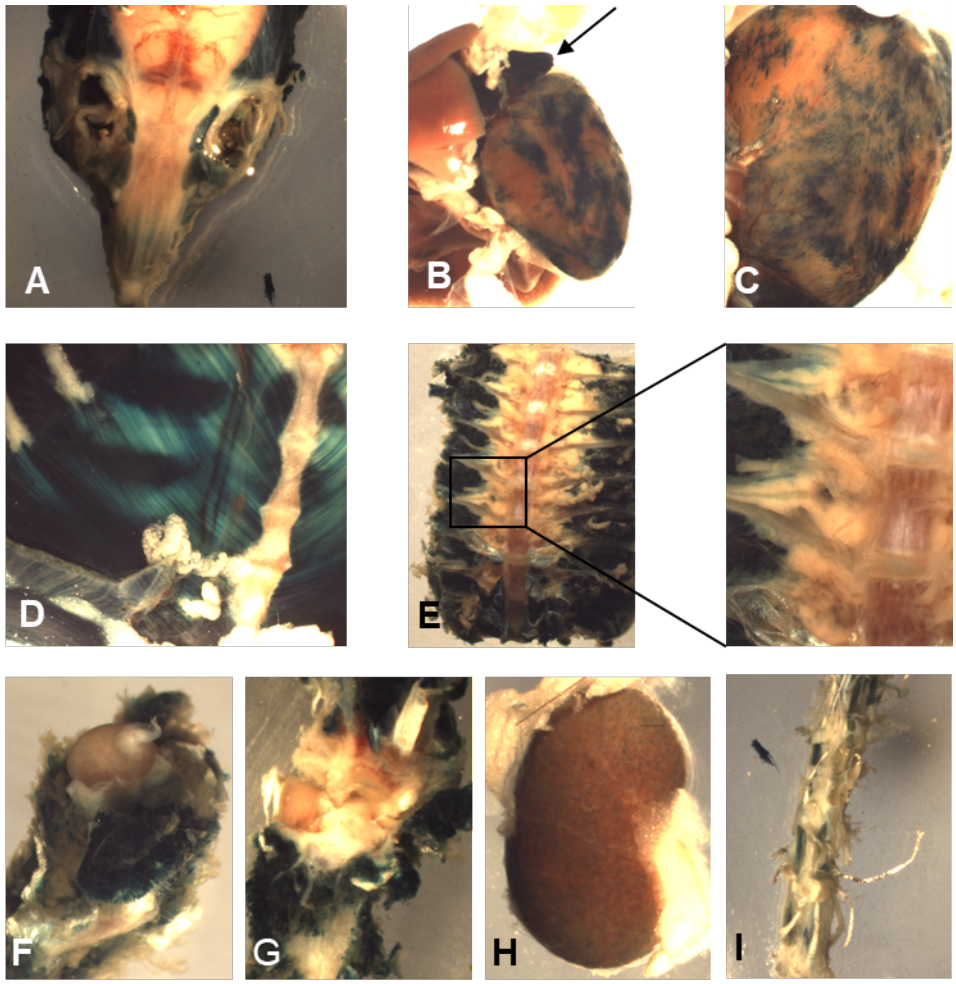


Figure 3.43: Whole mount images of X-gal stained tissue of founder C from -148kb_short line at six months of age. Intense staining was observed within the musculature surrounding the skeletal tissue of the cranium (A), ribs (D) spine (E), hip (F), knee (G) and tail (I). Staining also occurred in the heart, with strong staining of the atria (arrow, B), further staining in patched pattern occurred across the heart (B and C). Punctate staining of the kidney was also observed (H).

The pattern of transgene expression was reaffirmed upon histological sectioning (Figure 3.44).

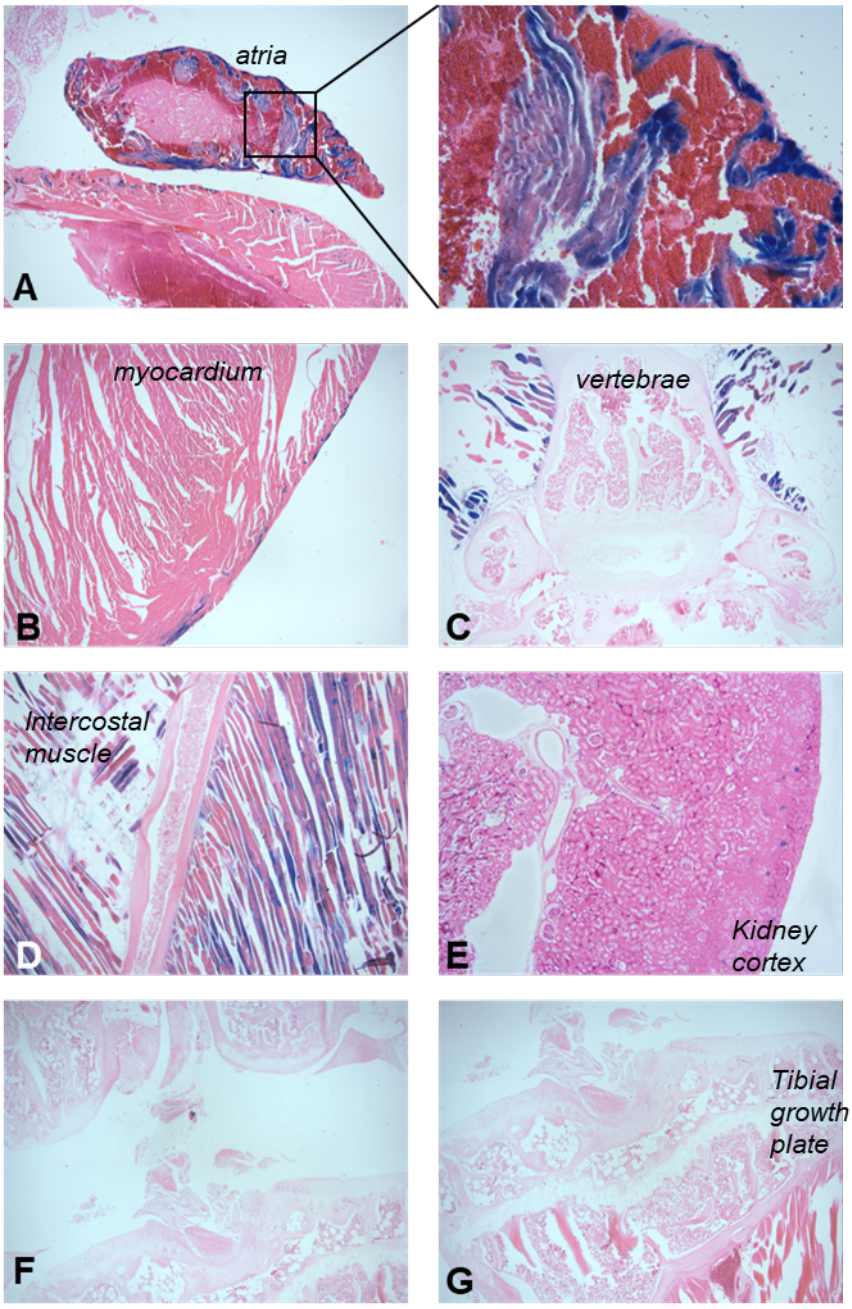


Figure 3.44: Histological sectioning of X-gal stained tissue of founder C of -148kb_short at six months of age. The most potent positive X-gal staining occurred within myocardium of the heart atria (A). Further cardiac staining was observed on the superficial surface of the heart (B), There was no staining of any cartilaginous or osseous structure such as within the vertebrae (C), ribs (D) the knee synovial cavity (F) or tibial growth plate (G). with There was rich staining of the musculature such within the intercostal muscles (C and D). There was further superficial staining in the kidney (E).

3.5.4 -148kb_short founder D exhibited transgene activity in mature adulthood

The activity of the β-galactosidase transgene product was assessed at mature adulthood (approximately six months) for -148kb_short D. Whole mount imaging

demonstrated striking staining co-localised with the vasculature across many tissues, as shown in Figure 3.45.

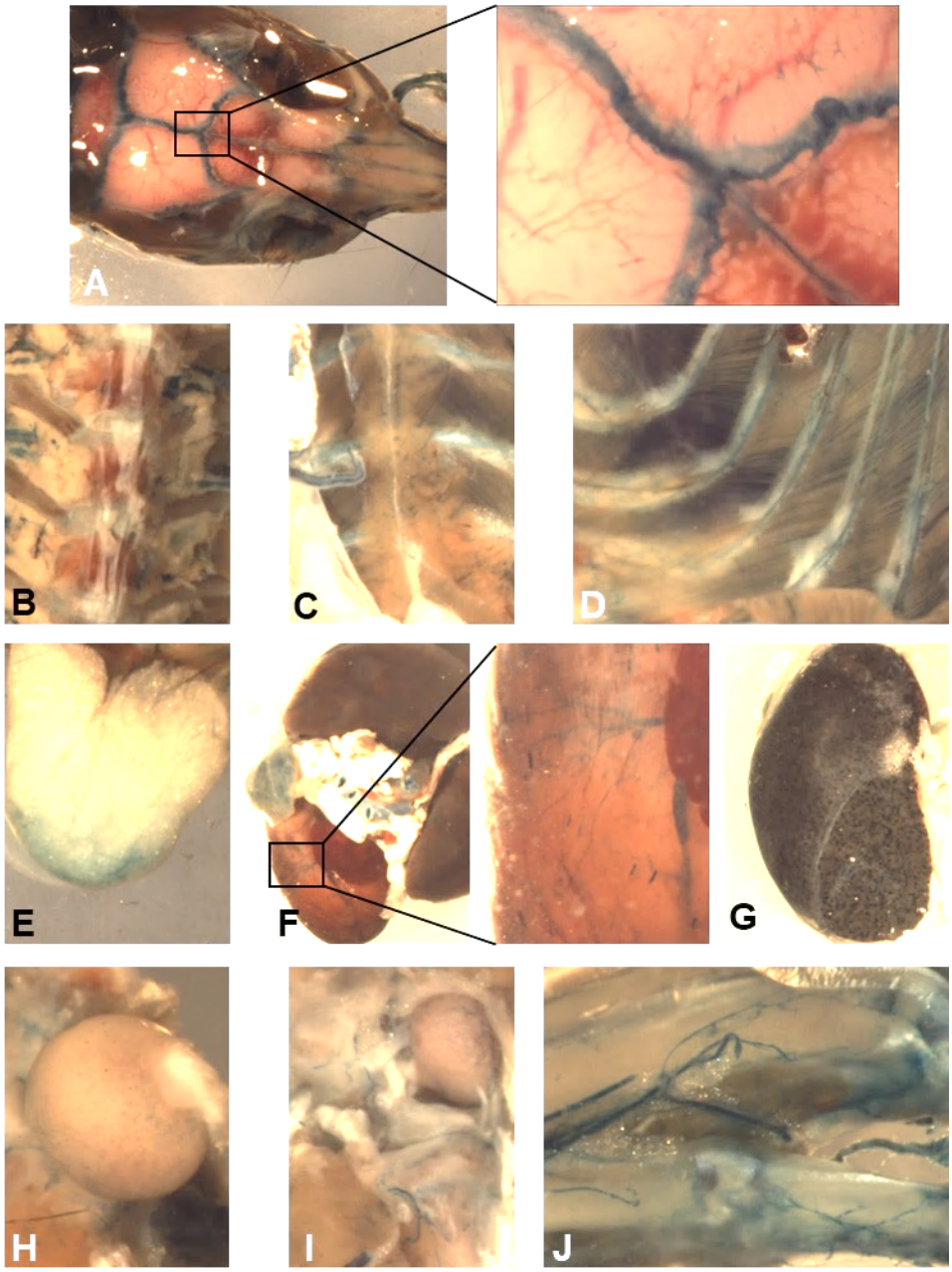


Figure 3.45: X-gal stained tissue from founder D of the -148kb_short transgenic line at approximately six months of age. Potent β-galactosidase activity was observed within several tissues Transgene activity within the musculoskeletal system was most striking within the cranial sutures (A) but was also observed; albeit in a weak manner, within the caudal portion of the xiphoid process (E) and articular surfaces of the hip (H) and knee (I).. The most striking blue staining was confined to blood vessels such as across the surface of the ribs (C and D) heart (F), kidney (G) and foot (J).

Further transgene activity was also recorded within musculoskeletal tissue, the most potent of which occurred in the cranial sutures (Figure 3.45 A). Very weak staining of the hyaline cartilage of the xiphoid process (Figure 3.45 E) and articular

cartilage of the femoral head (Figure 3.45 H) and femoral condyles (Figure 3.45 I) was also observed. This was visualised more effectively with histological sectioning of tissue (Figure 3.46).

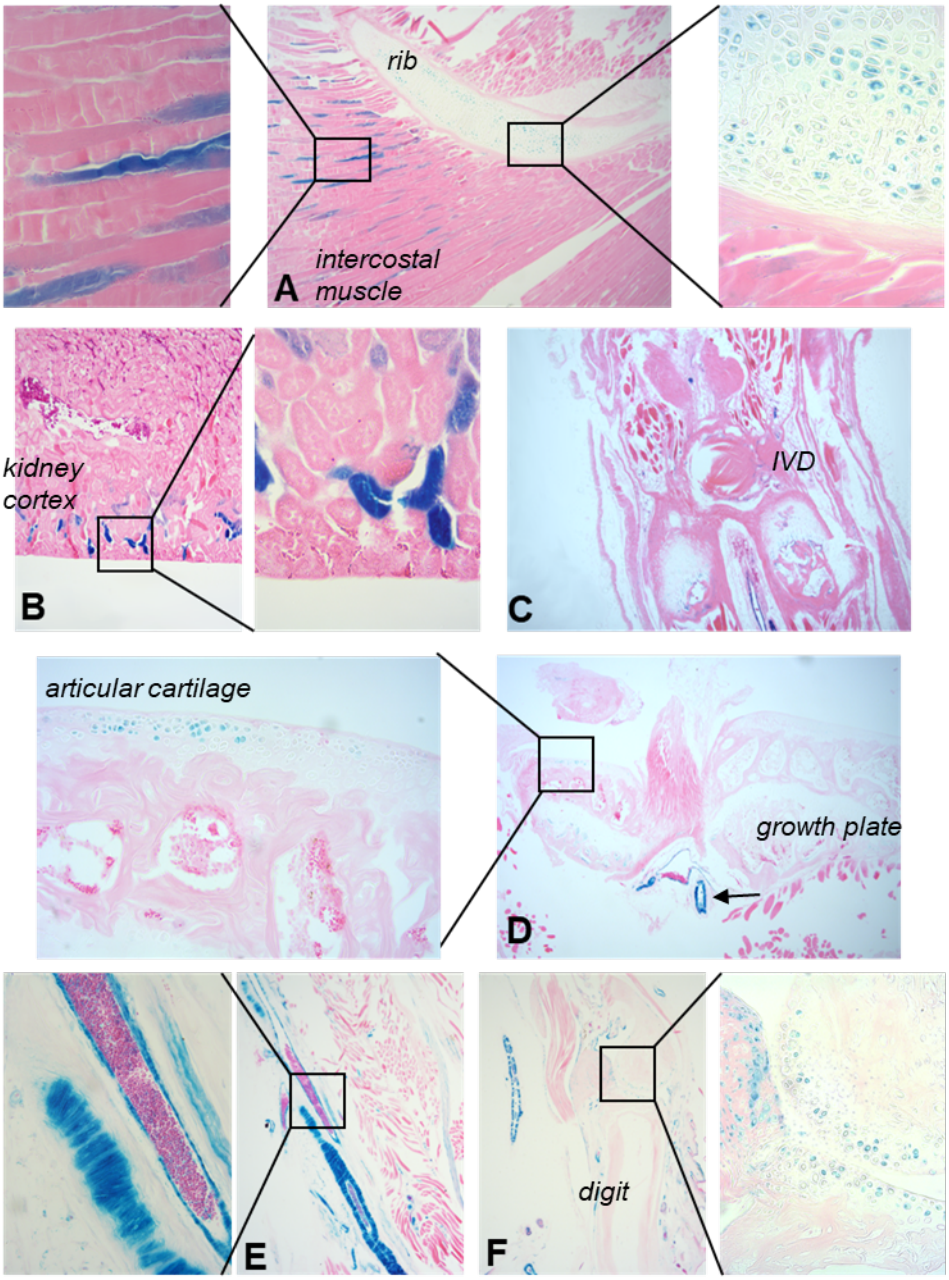


Figure 3.46: Representative images of histological sections of X-gal stained tissue from -148kb_short D in mature adulthood. Several tissues stained positively for β-galactosidase activity at approximately six months of age. Striated staining was observed within the intercostal muscles (A). Sparse staining was observed within hyaline cartilage including the costal cartilage (A), and articular cartilage of the vertebral endplate (C), tibial plateau of the knee (D), and hind paw digits (F).The most prevalent staining occurred within the vasculature globally, this was well demonstrated within the tail vein (C), popliteal artery of the knee (arrow, D) and digital arteries of the hind paw (E).

Striated staining of the costal musculature observed in whole mount imaging (Figure 3.45 D) was also observed in histological sections (Figure 3.46 A). The staining of the kidney visualised in Figure 3.46 G was found to occur in a superficial manner within the cortex of the kidney (Figure 3.46 B).Staining occurred within chondrocytes of hyaline cartilage in a global; albeit weak, manner. There was no staining of osseous cells. The vascular localisation of staining was clearly observed within sections globally, with erythrocytes present within richly stained vessel structures, such as within the popliteal artery at the tibial plateau (Figure 3.46 D) and paw (Figure 3.46 F). There was sparse staining of chondrocytes within hyaline cartilage, such as within the costal cartilage (Figure 3.46 A) and the articular cartilage of the tibial plateau of the knee (Figure 3.46 D) and digits (Figure 3.46 F).

3.6 Summary of enhancer driven reporter gene expression in vivo

The compilation of information regarding the reporter gene expression patterns allows comparisons to be made between each of the enhancers to drive cell lineage-specific transgene expression. For example -102kb functioned highly specifically and consistently, within the vasculature whereas -148kb functioned most reproducibly within chondrocytes. A summary of *in vivo* reporter gene assays is outlined in Table 3.9.

Enhancer	Timepoint	Founder	Tissue type						
			Cardiovascular		Musculoskeletal				
			heart	vasculature	Cartilage			Bone	Skeletal muscle
					Hypertrophic	Articular	proliferative		
-4kb	E15.5	A	No	No	No	No	No	No	No
		B	No	No	No	No	No	No	No
		C	No	No	No	No	No	No	No
-102kb	E15.5	1	No	Yes	No	No	No	No	No
		2	No	Yes	No	No	No	No	No
		3	No	Yes	No	No	No	No	No
		4	No	Yes	No	No	No	No	No
		5	No	Yes	No	No	No	No	No
		6	No	Yes	No	No	No	No	No
-137kb	E15.5	1	No	No	No	Yes	No	No	No
		1	No	No	No	Yes	No	No	No
		3	No	No	No	Yes	No	No	No
		4	No	No	No	Yes	No	No	No
		5	No	No	No	Yes	No	No	No
-148kb transient	E11.5	A	No	No	n/a	n/a	n/a	n/a	n/a
		B	No	No	n/a	n/a	n/a	n/a	n/a
		C	No	No	n/a	n/a	n/a	n/a	n/a
		D	No	No	n/a	n/a	n/a	n/a	n/a
		E	No	No	n/a	n/a	n/a	n/a	n/a
		F	No	No	n/a	n/a	n/a	n/a	n/a
		G	No	No	n/a	n/a	n/a	n/a	n/a
	E15.5	A	No	No	Yes	No	No	No	No
		B	No	No	Yes	No	No	No	No
		C	No	No	No	No	Yes	No	No
-148kb stable line	E11.5	n/a	Yes	Yes	n/a	n/a	n/a	n/a	n/a
	E12.5	n/a	Yes	Yes	n/a	n/a	Yes	n/a	No
	E13.5	n/a	Yes	Yes	Yes	Yes	Yes	n/a	No
	E15.5	-	No	No	Yes	Yes	Yes	No	No

	E17.5	-	No	No	Yes	Yes	Yes	No	No
	Nine weeks	-	No	No	Yes	Yes	Yes	No	No
	Five months	-	No	No	Yes	Yes	Yes	No	No
-148kb short	E15.5	A	No	Yes	No	No	Yes	No	No
		B	No	No	No	No	No	No	No
		C	No	No	No	No	No	Yes	No
		D	No	No	Yes	No	No	No	No
		E	No	No	No	No	No	No	No
	Five months	B	Yes	No	No	No	No	No	No
	Six months	C	Yes	No	No	No	No	No	Yes
		D	No	Yes	No	Yes	No	No	No
-198kb	E15.5	A	No	No	Yes	Yes	No	No	No
		B	No	No	No	No	No	No	No
		C	No	No	No	No	No	No	No
-230kb	E15.5	A	No	No	Yes	No	No	Yes	No
		B	No	No	No	No	No	No	No
		C	No	No	No	No	No	No	No
		D	No	No	No	No	No	No	No
	P7	C	No	No	No	No	No	Yes	Yes
	Six weeks	C	No	No	No	No	No	Yes	Yes
	Four months	C	No	No	No	No	No	No	No

Table 3.9: Summary of reporter gene expression patterns driven by each of the enhancers of Ccn2. Highly reproducible expression within vascular endothelial cells and articular chondrocytes was observed for the -102kb and -137kb enhancers. The -148kb enhancer drove function within chondrocytes across several founders. Founder variants of the -148kb_short line also demonstrated chondrocyte function, this was shown in several of the founders, however this expression was not as consistent as expression driven by the whole enhancer. Strong transgene activity was only observed in one of the three E15.5 -198kbHsp68LacZ embryos, limiting accuracy of transgene expression patterns driven by this enhancer.

3.7 In vitro assay of Ccn2 enhancer function

The capacity of putative enhancer regions to drive the expression of firefly luciferase reporter gene (*luc2*) was assayed in several cell types. Each enhancer region controlled the transcription of the reporter gene in conjunction with sequence corresponding to the *Ccn2* promoter region. Renilla luciferase gene expression was recorded for each sample and used to normalise data. This minimised error due to discrepancies in cell number or transfection efficiency. Results were gathered from three independent experiments for each cell type, where each experimental condition was repeated with three biological replicates. The error bars in each graph represent standard deviation

3.7.1 *In vitro* reporter gene expression within chondrocyte cell lines

The HTB94 and TC28-i2 human chondrosarcoma derived immortalised cell lines were used as models for the *in vitro* assay of putative enhancer function within chondrocyte like cells. Given the precedent set from findings of *in vivo* reporter gene expression that the -137kb, -148kb and -198kb enhancers each functioned within chondrocyte cells (Chapter 3.2-3.6), it was expected that strong luciferase reporter gene would be observed in cells transfected with vectors containing these regions. However, this was not observed in TC28-i2 or HTB94 cells (Figures 3.47 and 3.48 respectively).

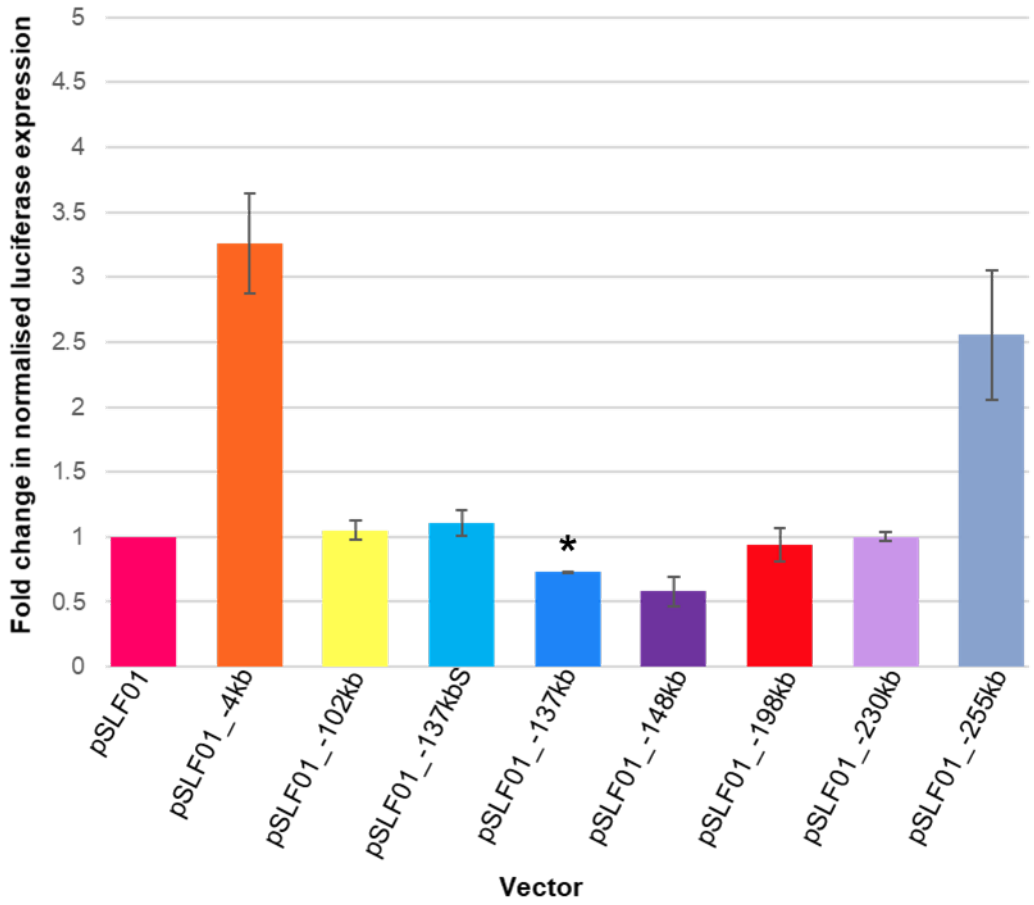


Figure 3.47: Fold change in normalised expression of luciferase driven by enhancers relative to *Ccn2* promoter in TC28-i2 cells, 24 hours after transfection. The largest change in luciferase production was observed for the -4kb and -255kb enhancers, however the average fold difference relative to the promoter was not significant due to variability in the data. There was a significant decrease in the amount of luciferase produced by cells transfected with the full -137kb enhancer relative to the promoter alone. Error bars represent standard deviation and * denotes significance $p \leq 0.05$

In TC28i2 cells transfected with the enhancer containing luciferase expression vectors, the -4kb and -255kb enhancers facilitated the greatest fold change in luciferase expression relative to the promoter alone, however neither of these results was statistically significant. Surprisingly, the -148kb enhancer, which demonstrated strong chondrocyte related function *in vivo* did not drive reporter gene expression. The reduction in luciferase activity controlled by this element was not statistically significant. Interestingly, there was a difference between the luciferase activity driven by the -137kb short and -137kb full enhancer variants. The increase in fold change relative to pSLF01 was not statistically significant for the shorter version of the enhancer, however the decrease of approximately 20% in luciferase activity regulated by pSLF01_137kb compared to the control was statistically significant (p 0.008). This difference would have been due to the presence of TFBS within the full -137kb that were utilised by repressive factors that

prevented enhancer function and therefore increased *luc2* expression. These results suggest that enhancer function did not have a significant impact on the transcriptional output of the *luc2* compared to that driven by the promoter region alone in the period 24 hours after transfection.

This experimental procedure was repeated in HTB94 chondrocytes. The trend of fold change in normalised luciferase expression relative to that controlled by the promoter alone (Figure 3.48) was similar to that observed for the TC28-i2 cells.

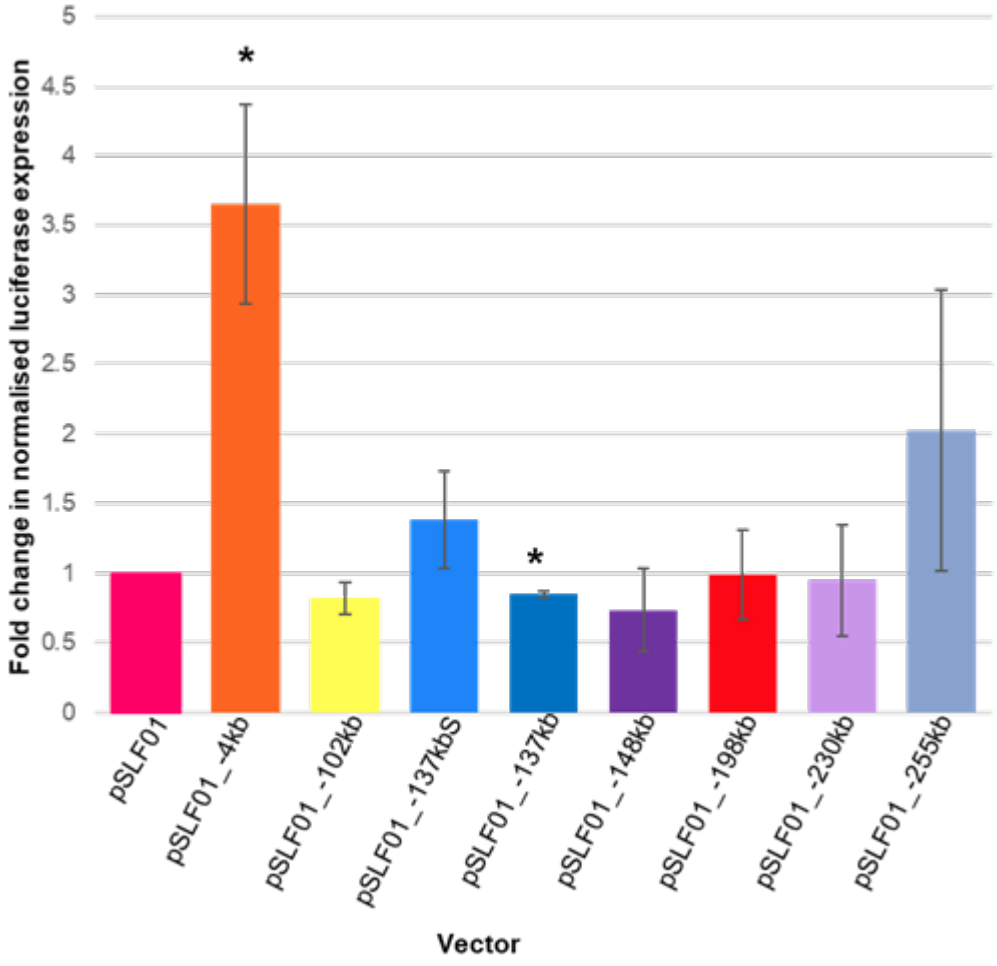


Figure 3.48: Fold change in normalised expression of luciferase driven by enhancers relative to *Ccn2* promoter in HTB-94 cells, 24 hours after transfection. The largest fold increase of approximately 3.5 times luciferase expression relative to the promoter alone occurred in cells treated with the -4kb enhancer, which was statistically significant P≤0.05. The small decrease in luciferase observed in cells treated with the -137kb full enhancer was also significantly significant. The results for -148kb, -198kb and -230kb were not statistically significant, with average similar to that of the promoter alone, suggesting negligible reproducible enhancer activity. Error bars represent standard deviation and * denotes significance p≤0.05

The -4kb and -255kb enhancers were observed to drive transgene expression with the largest fold changes of approximately 3.5 and 2 times respectively greater luciferase expression relative to the promoter alone. The result for the -4kb

enhancer was significant with a *P* value of 0.023 whereas the result for the -255kb was not. This suggests that transcription factors within the HTB-94 cells interacted with cognate TFBS within the -4kb sequence facilitating enhancer function and an increase in the expression of *luc2*. The result for the -137kb full enhancer was also significant, with a decrease in luciferase of approximately 15%, this echoed the data from the TC28-i2 cells with the -137kb short enhancer seemingly driving a small increase in luciferase expression. Again, this could be due to the presence of repressive elements within the full -137kb enhancer that are not present in the short region. As in the TC28-i2 cells, the -148kb region did not drive an increase in the expression of luciferase, which was surprising given the strong chondrocyte-related function of the enhancer *in vivo*.

3.7.2 *In vitro* enhancer expression in non-chondrocyte cells

The capacity of the enhancers to drive luciferase expression within pre-osteoblast cells was examined using undifferentiated MC3T3-E1 cells. This cell line is originally derived from new-born mouse calvaria. The results obtained from this cell line (Figure 3.49) were similar to those obtained in the chondrocyte like cells (Figures 3.47 and 3.48). The -4kb enhancer facilitated a significant fold change with approximately 3.5 times more luciferase expression relative to the action of the *Ccn2* promoter alone. The results for the -102kb, -137kb full, -148kb, -198kb and -230kb enhancers all constituted significant changes in fold change relative to the promoter alone, which each facilitating approximately 50% decrease in luciferase expression. This result was surprising for the -230kb enhancer given that it functioned in osteoblastic cells *in vivo* (Figures 3.11 to 3.17).

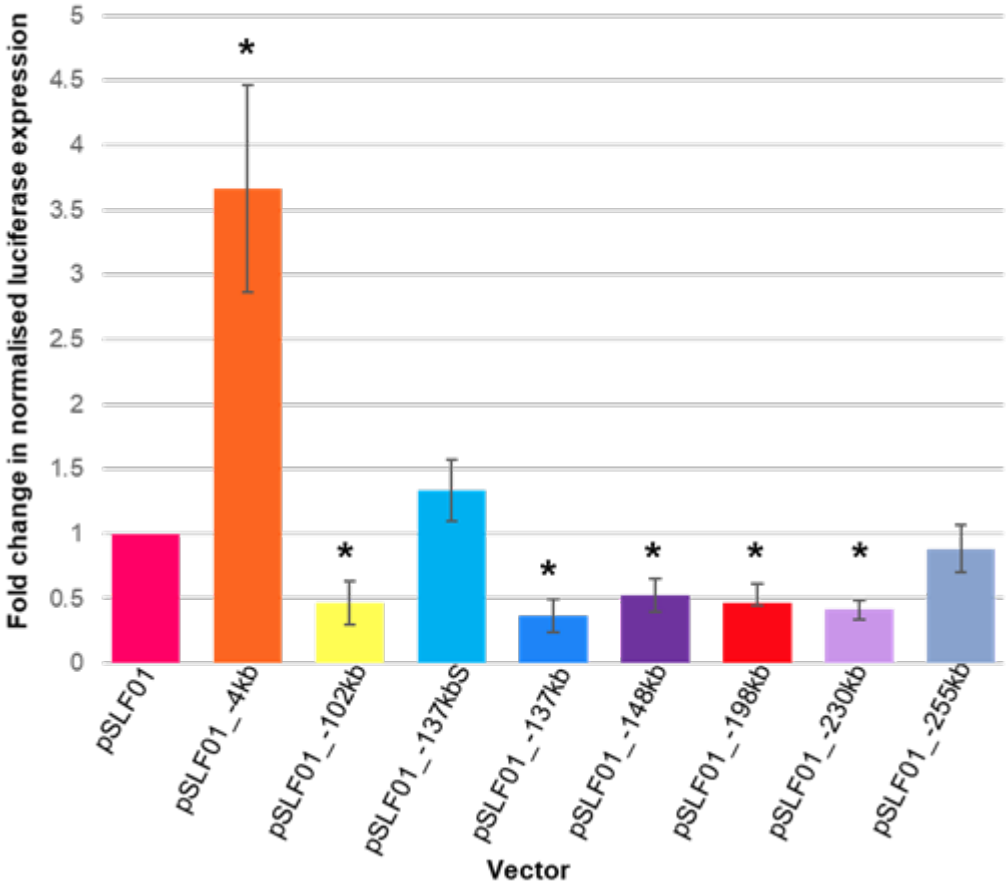


Figure 3.49: Fold change in normalised luciferase expression relative to the activity of the *Ccn2* promoter in MC3T3-E1 pre-osteoblast cells, 24 hours after transfection. A statistically significant increase in normalised relative luciferase expression was exhibited in cells transfected with the -4kb enhancer driving transgene expression. There were statistically significant (* denotes $p \leq 0.05$) decreases in normalised luciferase of approximately 50% in cells transfected with *luc2* gene under control of the -102kb, -137kb full, -148kb, -198kb and -230kb enhancers relative to the promoter alone. Error bars represent standard deviation.

Decrease in fold change in luciferase expression driven by the enhancers relative to the *Ccn2* promoter was also observed within murine NIH3T3 fibroblastic cells (Figure 3.50). The greatest reduction in reporter expression compared to the pSLF01 control occurred in cells transfected with the -137kb full enhancer with decrease of approximately 80%. As in the other cell types, the shorter version of this sequence facilitated a greater level of gene expression, further suggesting the presence of additional repressive sequences within the full enhancer that prevent the activity of the promoter and enhancer regions. Both the -198kb and -230kb enhancers were capable of driving transgene activity in fibroblasts *in vivo*, yet the action of neither of these regions led to an increase in luciferase activity within the NHI3T3 cells.

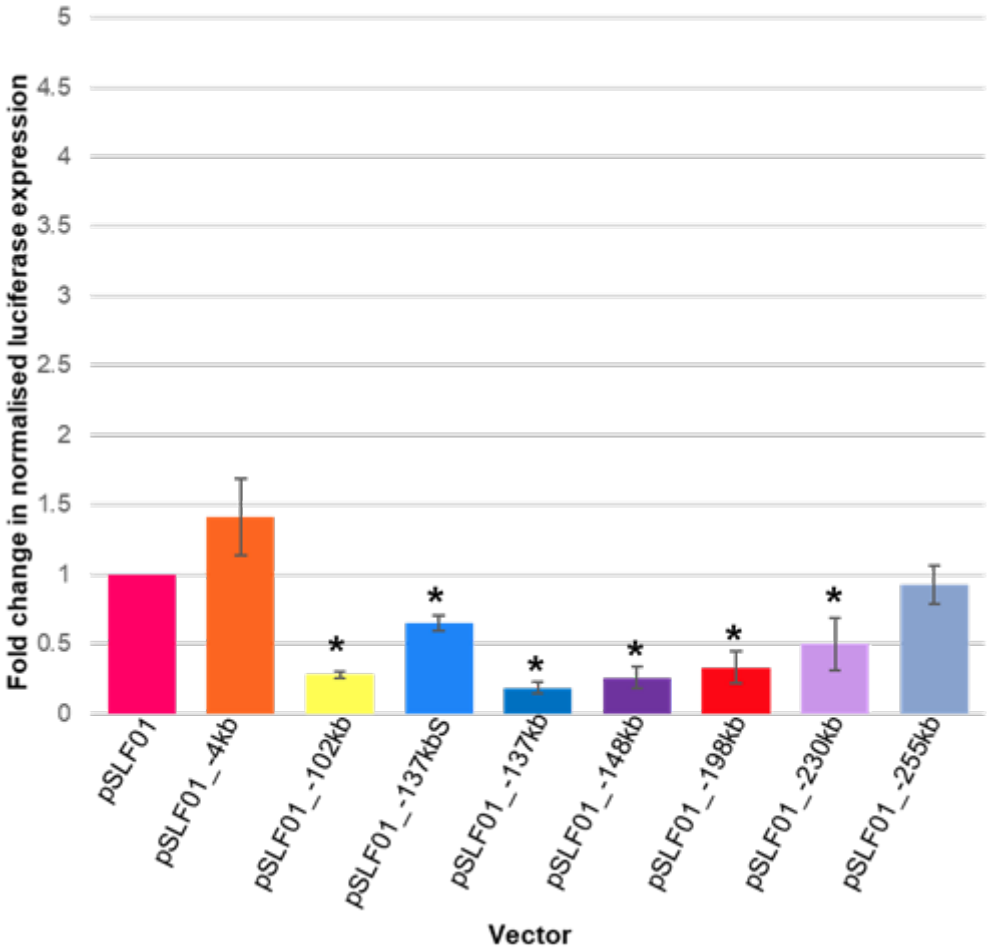


Figure 3.50: Fold change in normalised luciferase expression relative to the activity of the *Ccn2* promoter in NIH3T3 cells. There were significant decreases in normalised luciferase expression of at least 50% in cells transfected with the -102kb, -137kb short and full, -148kb, -198kb and -230kb enhancers, compared to pSLF01 promoter alone Error bars represent standard deviation and * denotes significance $p \leq 0.05$.

3.8 Identification of putative TFBS within Ccn2 enhancers

Given the tissue specific patterns of reporter gene expression driven by transgenic mice harbouring the enhancer regions, further examination of the sequences of the enhancers was conducted in order to try and understand the cellular signalling mechanisms that mediate regulation of enhancer function. This was given precedence over the enhancers that functioned well *in* vitro due to time constraints, and so the -137kb, -148kb and -230kb regions were chosen for further scrutiny and the identification of TFBS.

3.8.1 The -137kb enhancer contains a SOX9 consensus binding motif

Putative TFBS within the sequence of the -137kb enhancer were identified using publicly available ChIP-Seq datasets and the TRAP tool. The TRAP tool was used as it is publicly available and contains a comprehensive list of transcription factors and their predicted binding motifs. (Thomas-Chollier *et al.* 2011). Given the articular chondrocyte localisation of β-galactosidase expression in -137kbHsp68LacZ transgenic mice at E15.5, chondrocyte related genes were prioritised in the examination of TFBS.

SOX9 was principally examined due to its role as its fundamental role in chondrocyte physiology, as outlined in Chapter 1. Three ChIP-Seq datasets concerning SOX9 TFBS in chondrocytes were used. Firstly SOX9 in P1 costal chondrocytes and E17.5 nasal chondrocytes (GSM1692996 and GSM1693007 respectively; Ohba *et al*. (2015)). In addition to E12.5 limb chondrocytes (GSM1888973; Garside *et al*. (2015)). The peaks for SOX9 binding within -137kb lay within a sequence that is conserved to lizard (Figure 3.51). A 1275bp shorter region that contained the conserved sequence was chosen for further analysis (-137kb short, Figure 3.51), with co-ordinates chr10:24,178,225-24,179,500. When this sequence was inputted into the TRAP tool, SOX9 matrix M00410 ranked with 6th highest affinity for the sequence with probability of 0.00896 (Chapter 5.4). This motif has the consensus sequence NNNDAACAATRRNN.

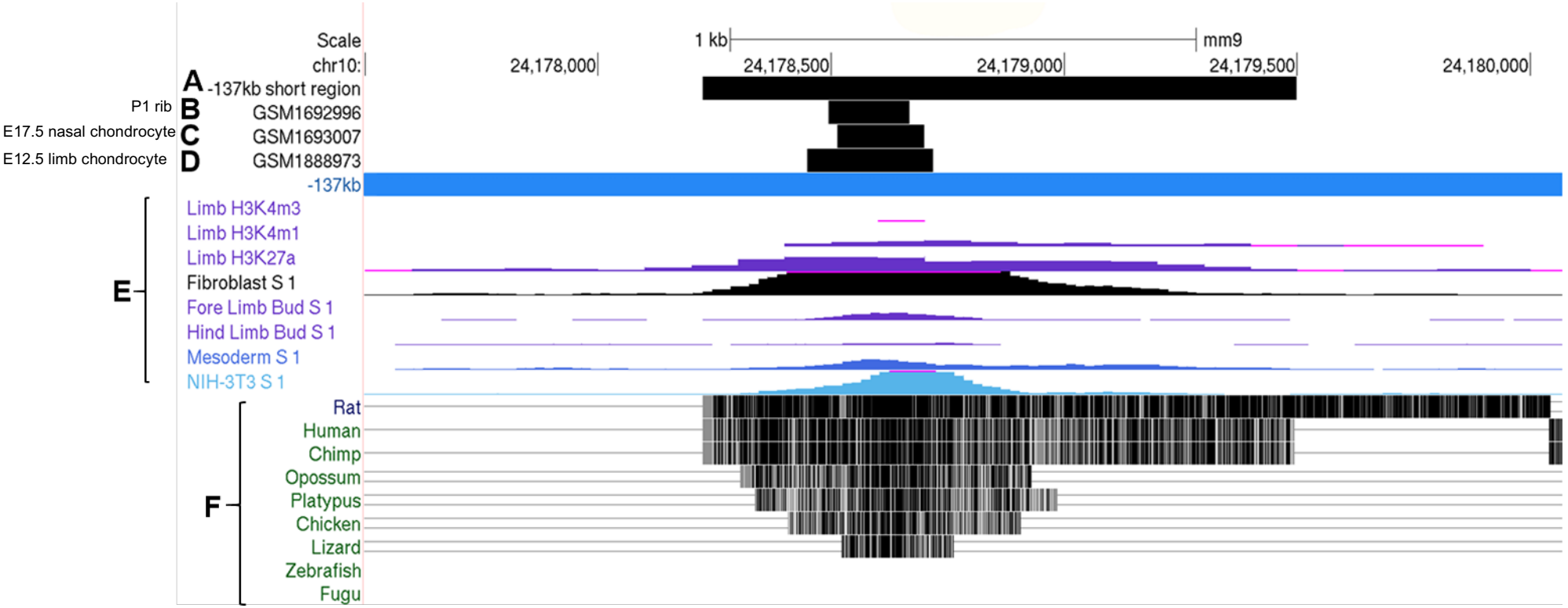


Figure 3.51: Identification of SOX9 binding sequences within the -137kb enhancer region using ChIP-Seq datasets. Typical enhancer chromatin characteristics of histone modification and DNase I hypersensitivity peaked in the middle of the region (E). This also concurred with the most conserved region of the sequence which is approximately 250bp of sequence highly conserved to lizard (F). Peaks for SOX9 ChIP-Seq datasets for P1 rib chondrocytes (GSM1692996, B), E17.5 nasal chondrocytes (GSM1693007, C) and E12.5 limb chondrocytes (GSM1888973, D) converged within the most conserved region of the sequence. 137kb short region (A) was selected based on conservation and the fact that histone modifications fell within this area

Combination of the SOX9 ChIP-Seq TFBS and TRAP predicted motifs allowed the prediction of SOX9 consensus binding motif located 346bp to 359bp in the -137kb short sequence (red and bold typeface in sequence and highlighted yellow in matrix information, Figure 3.52).

```
   1 acaaCAACAACAAAAGAAagaattcattaacaacagcaaagtcctccaccattatttttt
  61 tttctttccattaggtatgctcaaggctgtggcagaggaaatttctgcagtgttttgtt
 121 tcattttcttatgagaattactgatccaattgaatgtggcacctccttggttttaacccc
 181 tgtgctcttgctctagtcatctctgtcctttcacccccagaaacatcctgccaccactgg
 241 tttacttcagtttaattacaaacattttcattttctgatccatcaaga agtctagtcttc
 301 gctaagcatgcactgctcactccagacacataacacttaacaggcACTGAACAATGGGGg
 361 cctgttctctcctgtgaagcttccagaaacaggctttagggagctgccatgtgtttatct
 421 ttcttaacactccaacaggcaccctccatgactcacacacacagaaagccaggggcctt
 481 gagtggttgtttttttctctgcagctggacaaccctgcgcatggctctcagcaagggaga
 541 caagctttggggaggtgctcatgtagtcctgtagatcaaacacaaacaaaccctcaggag
 601 cttgcagtttccccacatcctttaagcaagtttagaattgacaagaaaatcagctcagca
 661 agccatctttacttcttctcacttaggtacaaatgagagaggacagtggtgaccttaaat
 721 ggcacaaatacaaagaagcttgaggtcatttccaagagtgtagaatcagtacttatcctg
 781 gagaagtgactctctatgtctatttaacataagactgctgaagacgtcgctacatcctct
 841 tgccctgcttgctttggaggatgaaacatgatgaatttccggtgagctctgagggtggcc
 901 cagctggtagctgccctggcatctggttgttctccccagtcttgtagagttttaacttat
 961 gagttaatattcagaagCAAGCACAAAGGGCctggtgtaaaaagactgaggtccttccct
1021 cagaacactatttttatcttcagatgtctctgcctgcctgaaccctaggctctagctct
1081 tctactctgcaagactacagggaGCCTATTGGTGTTCttgctctccgtgcatggtgttta
1141 agtgcctttatgtatcaagcaaggacattcactgggctcacataaataaatgccctcctt
1201 ttgtacaacatagatatgactattattatttctgtaagcctccatctagatttctagcag
1261 tattgtgtgactct
```

Individual sites (matrix, location, and weight score)

```
M00410 5-18(+) 6.0155
M00410 346-359(+) 10.3185
M00410 978-991(+) 6.14416
M00410 1104-1117(-) 5.01187
```

Figure 3.52: TRAP prediction of SOX9 binding motifs within the -137kb short region. The sequence for -137kb short was inputted into the TRAP tool, with red text highlighting four predicted SOX9 binding motifs based on the M00410 SOX9 consensus motif; NNNDAACAATRRNN. Sequence where all three SOX9 ChIP-Seq peaks from Figure 3.51 overlap is highlighted with bold typeface. One of the predicted SOX9 binding motifs sat within the ChIP-Seq interaction region (highlighted yellow).

Evolutionary conservation was then examined for the most conserved region of sequence within -137kb short; approximately 250bp that was conserved to lizard (Figure 3.51). DNA sequence was taken from the Multiz 30-way alignment tool in the UCSC Genome Browser for mouse, human, opossum, chicken and lizard and aligned using CLUSTALW (Figure 3.53). The strongest candidate for SOX9 binding motif, as aforementioned, was highly conserved between all organisms with the CAA motif that SOX9 interaction depends on replete in each organism. This motif and surrounding sequence was chosen to create an oligonucleotide probe for *in*

vitro validation of DNA sequence-SOX9 interaction (highlighted yellow, Figure 3.53).

```
CLUSTAL 2.1 multiple sequence alignment
lizard          TCTCTCC-ACATCTATTCCTCACTCCAGATGCATATT--TTTTCAAACTCAGAACAATGG
chicken         TCTCTCCTGCTTCCATTCCTCTCTCCAGCCTTGTAATACTCTGCAGACTCTGAACAATGG
opossum         ----------TCCATGGCCCACTCCAGCCCCATGAGCCTCTGCAGGCTCAGAACAATGG
human           -TCACTAAGCATGCATTTCCCACTCCAGACTCCAAACACTGTACAGGCTTTGAACAATGG
mouse           -TCGCTAAGCATGCACTGCTCACTCCAGACACATAACACTTAACAGGCACTGAACAATGG
                          *  *   * * ******          *   **  *   *********
lizard          ATTAGCTGTTCTCTCCAAGTA-GGTTTCCAGAAACAGGCTTTAGAGTTTTGCCATATGTT
chicken         ATTAGCTGTTCCCTCCAGTTAAGGTTTCCAGAAATAGTCTTTAGAGTTTTGCCGTATGTT
opossum         -CTATCTGTTCCCTCCTGTGA-GATTTCCAGAAACAGGCTTTAGAGCTCTGCCCTATGTT
human           -GTATCTGTTCCCTCCTGTCA-GGTTTCCAGAAATAGGCTTTAGAGAGCTGCCATGTGTT
mouse           -GGGCCTGTTCTCTCCTGTGA-AGCTTCCAGAAACAGGCTTTAGGGAGCTGCCATGTGTT
                   ****** ****    *    ********* ** ****** *   **** * ****
lizard          TATCTTTCTTAGTATTCCAGTGAGTAGCTCCTTTATAACTTTG-CATGCAAATAGCACC-
chicken         TATCTTTCTTAATACTCCAGTGAACAGCCCCCTCATGACTCAC-CG-GCAAACA------
opossum         TAGTTTTCTTCATACTCCCGCAAGCAGCCCCTCCATGACTCACACATACACACAAAGCCA
human           TCTCTTTCTTTACACTCCAGCAAGAA-CCCCTCCGTGACTTACATACACAGAAAG--CCA
mouse           TATCTTTCTTAACACTCCAACAGGCA-CCCCTCCATGACTCACACACACAGAAAG--CCA
                *   ******   * ***       * * **    * ***        ** * *
lizard          AAAATACTGAAAGTACTTGTTTGTTTGTTTTTACTCCTGAGGTGGATCATGTTGCAC-GT
chicken         AAGGTCTTGAATG---------GTTTGTTTATTCTTCTGAGCTAGATGGCACTGCCT-GG
opossum         AAGGCCCTGAATG---------GTTTGTTTATTCTCCCGGGCTACATGGTCCTGGACTGG
human           AAGGCCCTGAATG---------GTTATTTATTCTTGC-TAGCTGGATGACCCTGCAC-GT
mouse           GGGGCCTTGAGTG---------GTTGTTTTTTTCTCTGCAGCTGGACAACCCTGCGC-AT
                   ***  *         ***  **  *  *     * *  *      **
lizard          GAGTCACTGTGTGG
chicken         GGCTCACCACGTGG
opossum         GGTTCTCAGTGCGG
human           GGCTCTCAGCATGG
mouse           GGCTCTCAGCAAGG
                *  ** *     **
```

Figure 3.53: ClustalW alignment of the most evolutionarily conserved segment of the -137kb region. The mouse, human, opossum, chicken and lizard DNA sequences were aligned using ClustalW. The SOX9 consensus motif identified in Figure 3.50 is highlighted in green, with conservation (*) of the key CAA nucleotides within the motif between all organisms. The EMSA probe sequence which was used for *in vitro* validation of SOX9 TFBS is highlighted in yellow.

EMSA methodology was used to determine the capacity for the identified monomeric SOX9 consensus motif to interact with SOX9 protein. As previously mentioned a wild-type oligonucleotide probe was designed to span the identified SOX9 binding site with additional nucleotides to a length of 30nt (highlighted yellow Figure 3.53). Mutated oligonucleotide probes were also designed with substitution of A nucleotides for G in the SOX9 CAA recognition motif within the sequence. In these experiments, the wild-type oligonucleotide probe was bound by SOX9 protein (Figure 3.54).

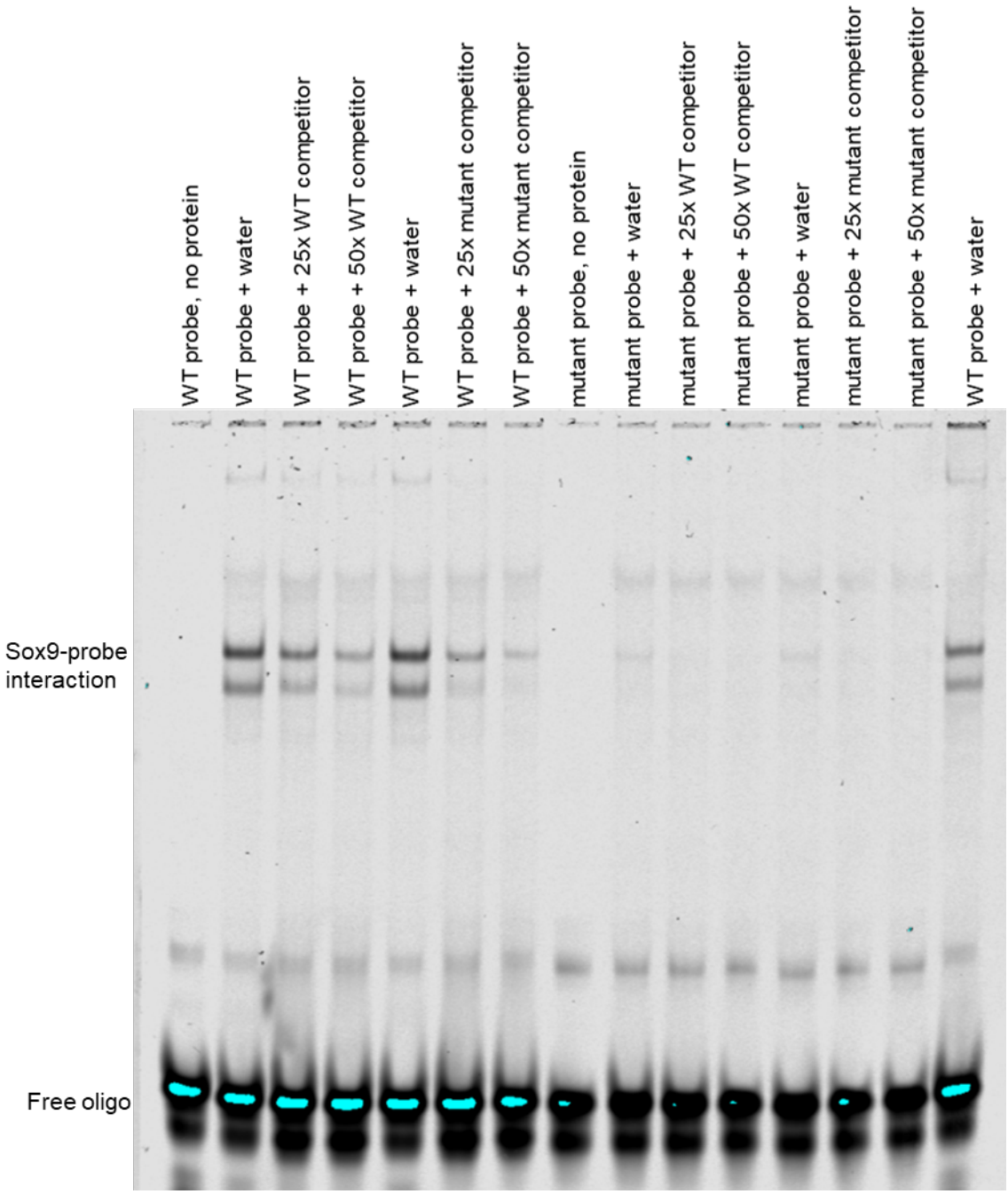


Figure 3.54: EMSA determination of interaction between *in vitro* produced SOX9 protein and consensus binding site within the -137kb enhancer DNA sequence. There was a shift in the signal produced by the probe in reactions containing the wild-type probe and SOX9 protein, representing DNA-protein interaction, with the migration of the probe signal being slowed by protein interaction. The intensity of this band was reduced with the use of wild-type competitor oligonucleotide, but not with a mutated competitor. There was no band present for interaction between the mutant probe and SOX9 protein.

There was a reduction in the intensity of the band in lanes representing reactions in which the wild-type competitor was used. This band was lost in the absence of SOX9 protein. Moreover, this band was lost with abrogation of the clear CAA site in the mutated oligonucleotide. However, the mutation of the oligonucleotide probe did not remove a sub-optimal Sox9 binding site, which therefore did not prevent

SOX9-TFBS binding. These observations demonstrate the specificity of interaction between the wild-type oligonucleotide and SOX9 protein.

3.8.2 The -148kb enhancer contains two SOX9 binding motifs

Given the strong, chondrocyte based reporter gene expression driven by -148kb *in vivo* from E15.5 in the stable transgenic mouse line, the sequence of the enhancer was examined in an attempt to identify master chondrocyte regulator SOX9 TFBS. Using the TRAP tool, SOX9 was not predicted to bind within the enhancer (Chapter 5.5). However, integration of ChIP-Seq datasets for SOX9 enabled visualisation of SOX9 TFBS in chondrocytes (Figure 3.55).

Within the enhancer, one sequence contained peaks for interaction in both P1 rib chondrocytes and E17.5 nasal chondrocytes (Ohba *et al.* 2015) with conservation of sequence to platypus. Sequence conservation within this sub-region was examined using ClustalW (Figure 3.56).

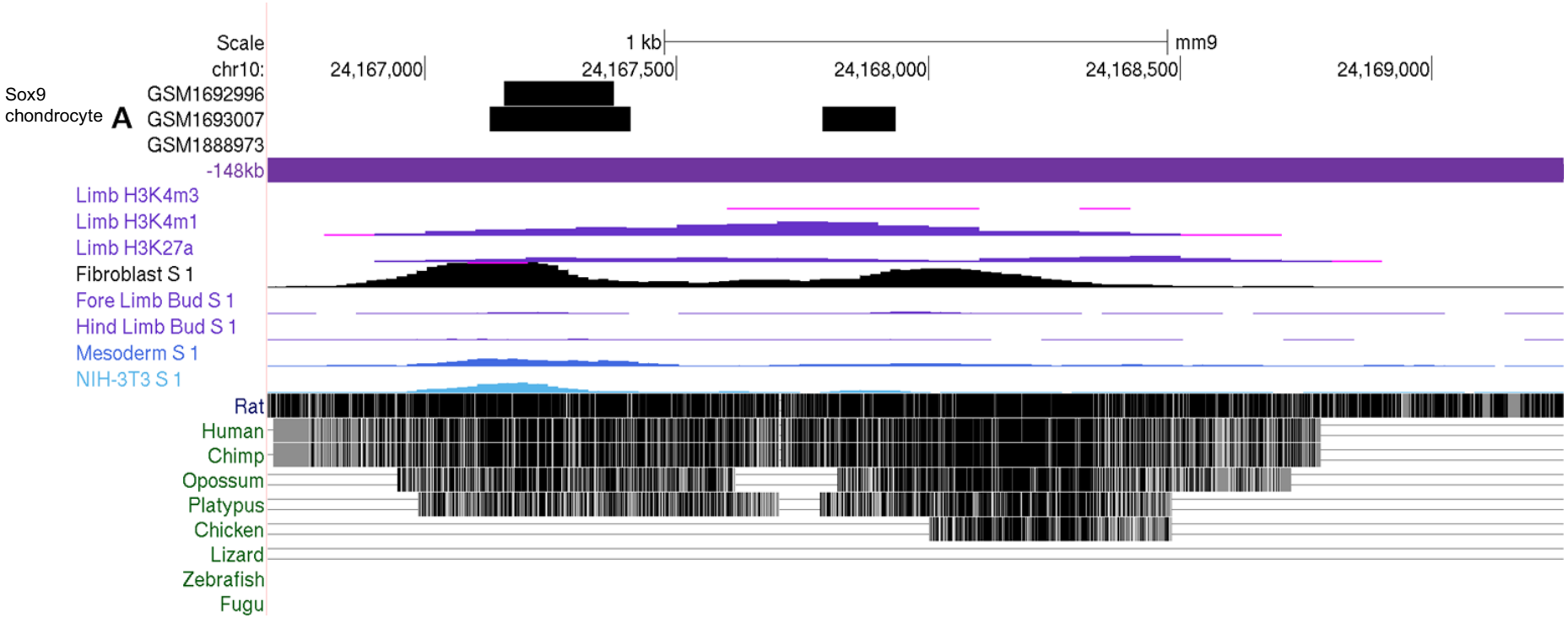


Figure 3.55: Identification of SOX9 binding sequences within the -148kb enhancer region using ChIP-Seq datasets. ChIP-Seq peaks for SOX9 TFBS (A, Ohba et.al (2015)) occur in two loci across the enhancer, with peaks being present in both E17.5 nasal chondrocyte and P1 rib chondrocyte in the 5' end of the enhancer, but only nasal chondrocytes for the 3' peak. Probes were designed in the 5' peak region that was not present in the -148kb_short transgenic mouse line.

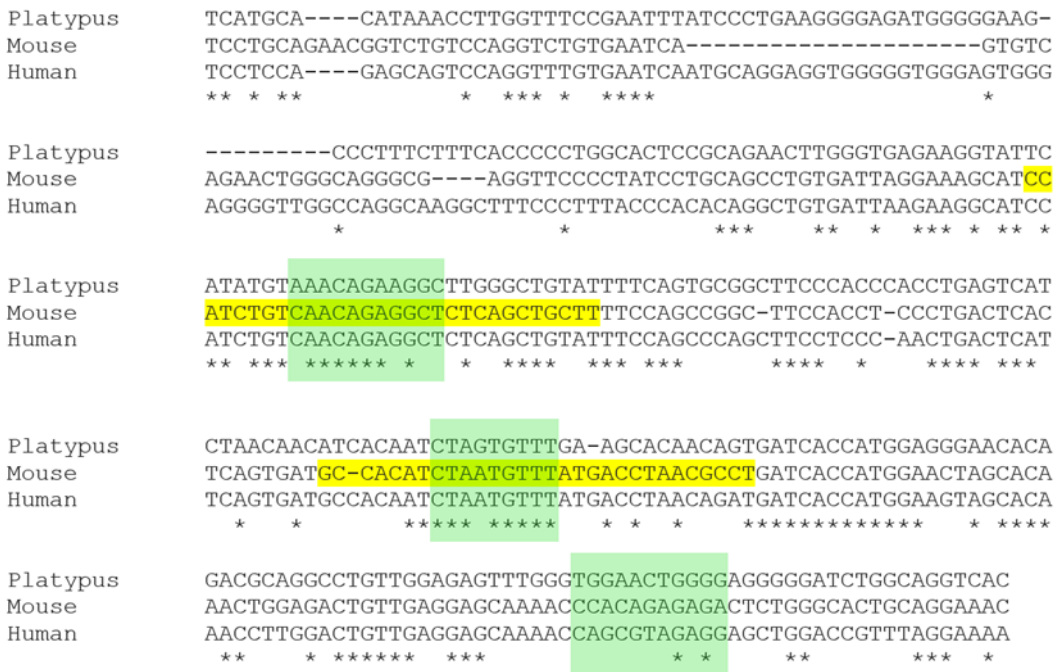


Figure 3.56: ClustalW alignment of sequence within a ChIP-Seq SOX9 peak in the -148kb sequence. There were three SOX9 motifs within the sequence and each are highlighted in green. The core motif of the first two motifs was evolutionarily conserved, however this was not true for the third motif. EMSA probes were designed to span the first two motifs, with sequences highlighted in yellow

Two conserved SOX9 binding motifs were chosen for *in vitro* examination of DNA-TF interaction in EMSA (sequences highlighted yellow, Figure 3.56). Both of these sequences lie within the sequence of the enhancer that was absent in the -148kb_short transgenic mouse line. EMSA methodology demonstrated that both of these TFBS are bound by SOX9 (Figure 3.57). A band was present at a higher position in the gel relative to the unbound free oligonucleotide, representing impaired probe migration through the gel due to interaction with the SOX9 protein. The interaction between the probes and SOX9 was specific, as a 25x and 50x excess of competitor oligonucleotides reduced the signal of these shifted bands. The use of a mutant competitor sequence also reduced shifted band signal although this effect was not as strong as in the wild-type competitor lanes. This suggests that mutation of the sequence did not completely prevent competitor oligonucleotide DNA-protein interaction. Sub-optimal binding of Sox9 could therefore occur without complete ablation of CAA consensus motif.

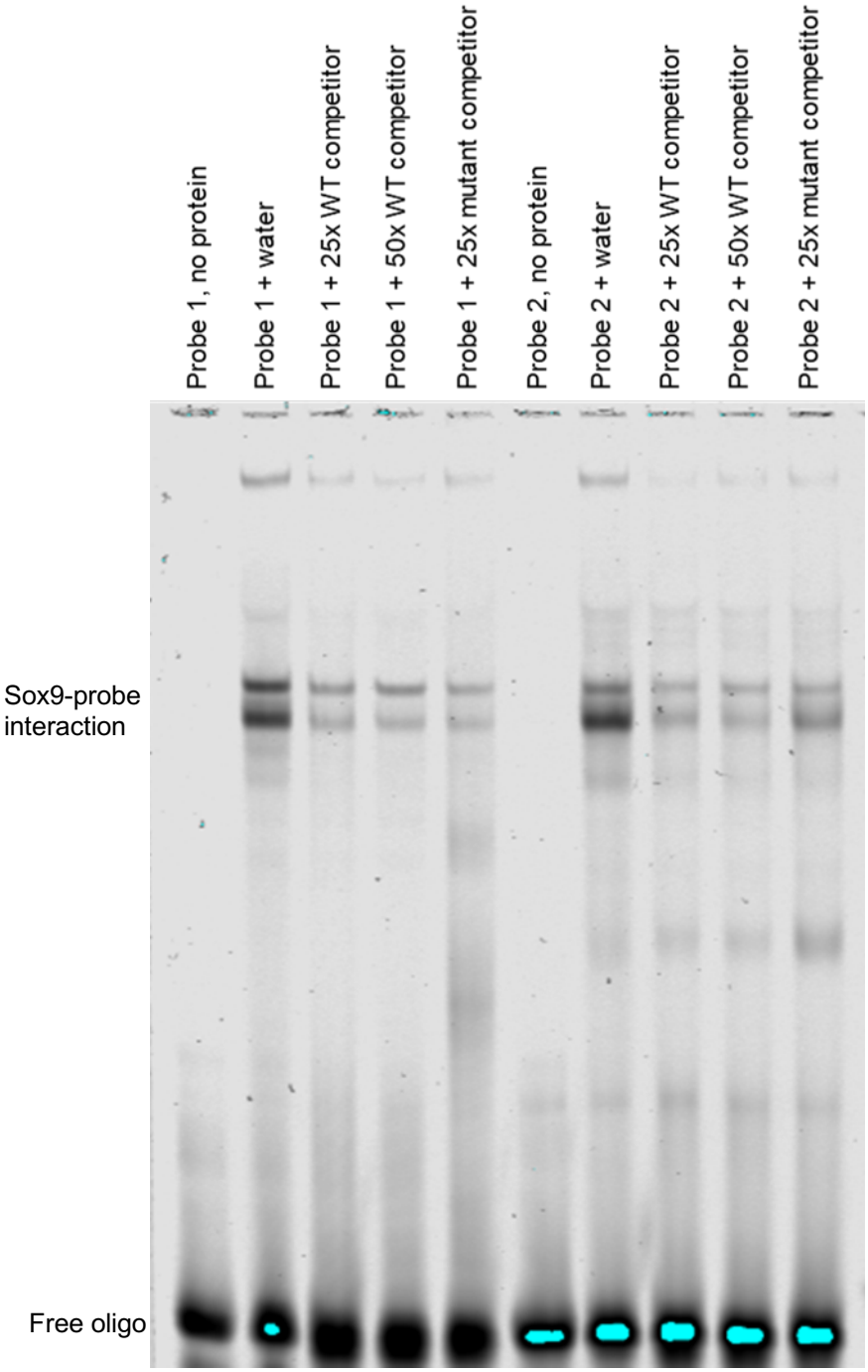


Figure 3.57: EMSA examination of -148kb sequence- SOX9 interaction. Two probes were used containing SOX9 consensus motifs. A shift in signal was observed with incubation of each probe and SOX9 protein, corresponding to DNA-TF interaction. Moreover, this interaction was specific as the use of wild-type competitor oligonucleotide greatly reduced the intensity the shifted band compared to reactions containing water.

3.8.3 The -230kb enhancer contains two SOX9 binding motifs

The process of identifying SOX9 motifs was repeated for the -230kb enhancer region. Firstly, the aforementioned SOX9 ChIP-Seq dataset peaks were compiled within the whole -230kb enhancer (Figure 3.58).

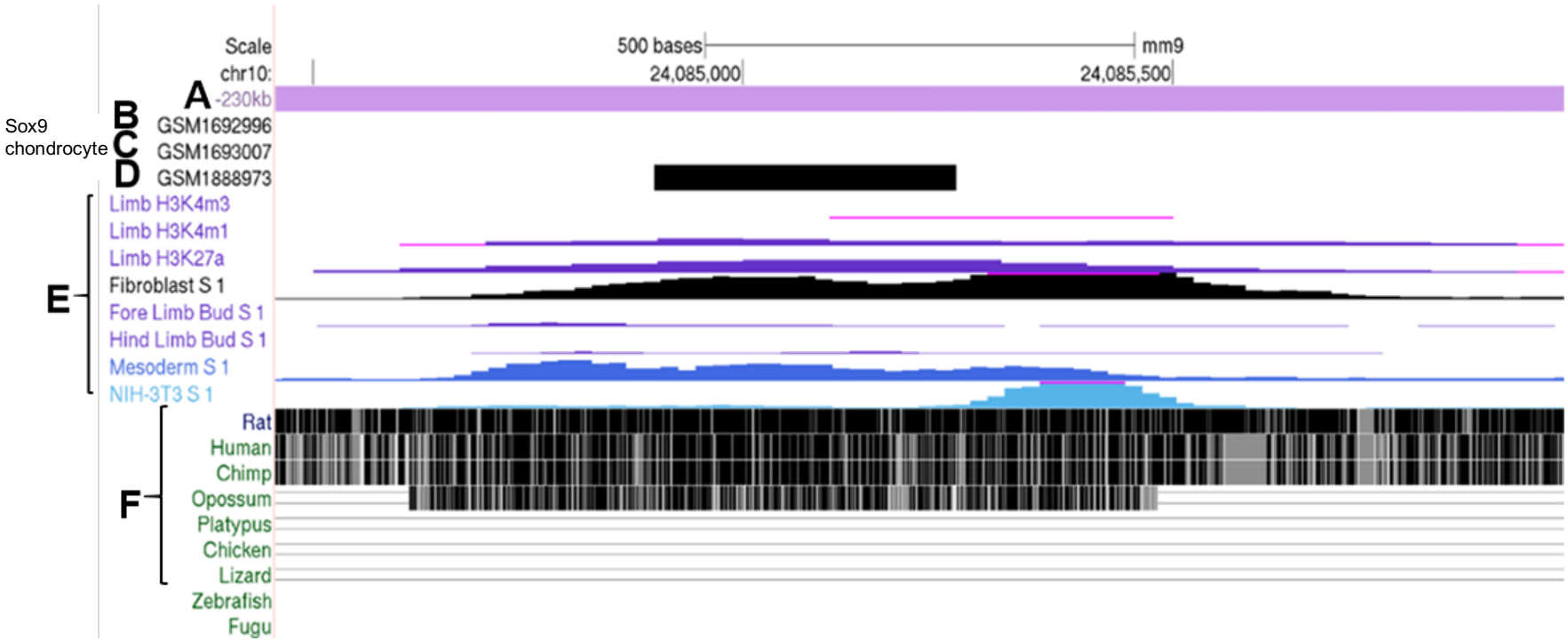


Figure 3.58: UCSC visualisation of SOX9 ChIP-Seq dataset TFBS within the -230kb enhancer region. Within the enhancer (A), there was a ChIP-Seq peak for Sox9 in one of the datasets (B-D). This peak lay within the region of the enhancer enriched for H3K4me1and H3K27ac in addition to DNase I hypersensitivity (E)). The most conserved fragment of sequence was shared between mouse and opossum (F). Peaks for SOX9 ChIP-Seq datasets only occurred in E12.5 limb chondrocytes (D), and not E17.5 nasal chondrocytes (B) or P1 rib chondrocytes (C)

This sequence was input into the TRAP and the M00410 SOX9 matrix was ranked with the 16th greatest affinity (Chapter 5.5). Within the sequence (Figure 3.58), none of the predicted binding sites (capitalised red typeface, Figure 3.59) for SOX9 sat within the peak for SOX9 binding within the GSM1888973 ChIP-Seq dataset (bold typeface, Figure 3.59) (Garside *et al.* 2015). An inverted repeat consensus Sox site was however manually identified within the ChIP-Seq dataset peak sequence, with a sequence of GACAAAGtacCTTTGTC (highlighted yellow, Figure 3.59).

```
   1 cattgagacataatcacattatgtatgcagggcaattttaacaaggctgagtaggcaaaa
  61 atcacactataatcactagtttaaaAAAAAACAAAAACAAAAACAAAATGAaaaagaaca
 121 aatcaactcttccggatgttacctgtaacaaattgtctcttttaactgtcctctggtggg
 181 cttatggtaactgtatcaagcaggaaagcagattcctaatggagttcttaaaacctgctg
 241 tgtaaacagtctcagaaactaaagaaaacagttcatgaatatagggacctgattcttcct
 301 ctttccccagccccagggaattaaggtctcacataacacaggttgcacttgctgtttgct
 361 gctacagcagaagttttctttgtaagccttctgcaagcaacattccgaaggggaagggag
 421 ctcactggtgttcagaactctgttgtttttctccctagatggttcaaatatattaacttg
 481 ggtttcaactcctccttcccacagaaagagcaggggaaagtgcttgtttttcttctggcct
 541 gggctgcgatgttttcctttcagctacagagaagtctagagggcatatgtctgggtcaga
 601 gacagatgtcagctgtgttttcagtatgcaggtaggtggtctcagactccttggcaaagt
 661 agccttataatgaaatgactgcagggacaaagtacctttgtcaacgtgccatcacaagct
 721 gtgctgctgtccccactgatagacttatcttaaaggttacagaccaccagaaaaactcaa
 781 actctgcacagacagtcgccttcaatgaatctccctttgaaacataaaataaagatcctc
 841 gaggcatgcgcagaccacatcagttatgcacagaccacatttgagtcgcataaatacctt
 901 gtgagcaacacaagaaccaCCTCATTGACCCAAagccatggtggacagctgccaagagac
 961 tcacagtgcaaccacagaaaatatatctaaaaccctagaaacaactgagagtaagtgctg
1021 ccaatgaagccatggcaacctCGTGATCAATTAAAtcaagtagaccatacttctttgcag
1081 aaacatcttagtttatattacctctgtcctttatgtctattatttttatctcgatcacgt
1141 tctccctgcctaactatataagcacattcctaattttgcagtctggttattgacctgtta
1201 aggcttccatgaaggtacttgggagtcatagatagttctacaggCAACAACAATTGATTA
1261 GTgtttgtggtttgtgagtgtgcgcgtgcgcgcacatATCTATTGTTCATActgattact
1321 ttagctccatccttctgtacttctttcatagacatcaaaaagaagcaggaagttaaacag
1381 acactgaatgtccagttctttaggaactgtgtgatcattaaaatttataaaagaactgac
1441 tgagaacctgagattcacatgtgagagatttccatttccaaagcctaggaagggagacg
```

Individual sites (matrix, location, weight score)

```
M00410 86-99(+) 5.45016
M00410 92-105(+) 5.29797
M00410 98-111(+) 5.20949
M00410 920-933(-) 5.00425
M00410 1042-1055(+) 5.53017
M00410 1245-1258(+) 7.17254
M00410 1249-1262(-) 5.63419
M00410 1298-1311(-) 9.87402
```

Figure 3.59: Visualisation of TRAP predicted SOX9 binding sites within the -230kb enhancer. None of the SOX9 M00410 motifs (red typeface) sat within the ChIP-Seq peak area (bold typeface). However an inverted repeat SOX9 motif was identified within this region (yellow highlight).

The most conserved sub-region of -230kb is approximately 900bp that exhibits conservation to opossum in the Multiz tool. The sequences for mouse, human and opossum were aligned using ClustalW (Figure 3.60). One TRAP predicted

sequence fell within this sequence, but was not highly conserved between the species (red typeface, Figure 3.60). The core nucleotides of the inverted repeat motif GACAAAGtacCTTTGTC (blue, bold and underlined typeface Figure 3.60) identified in the ChIP dataset peak were conserved.

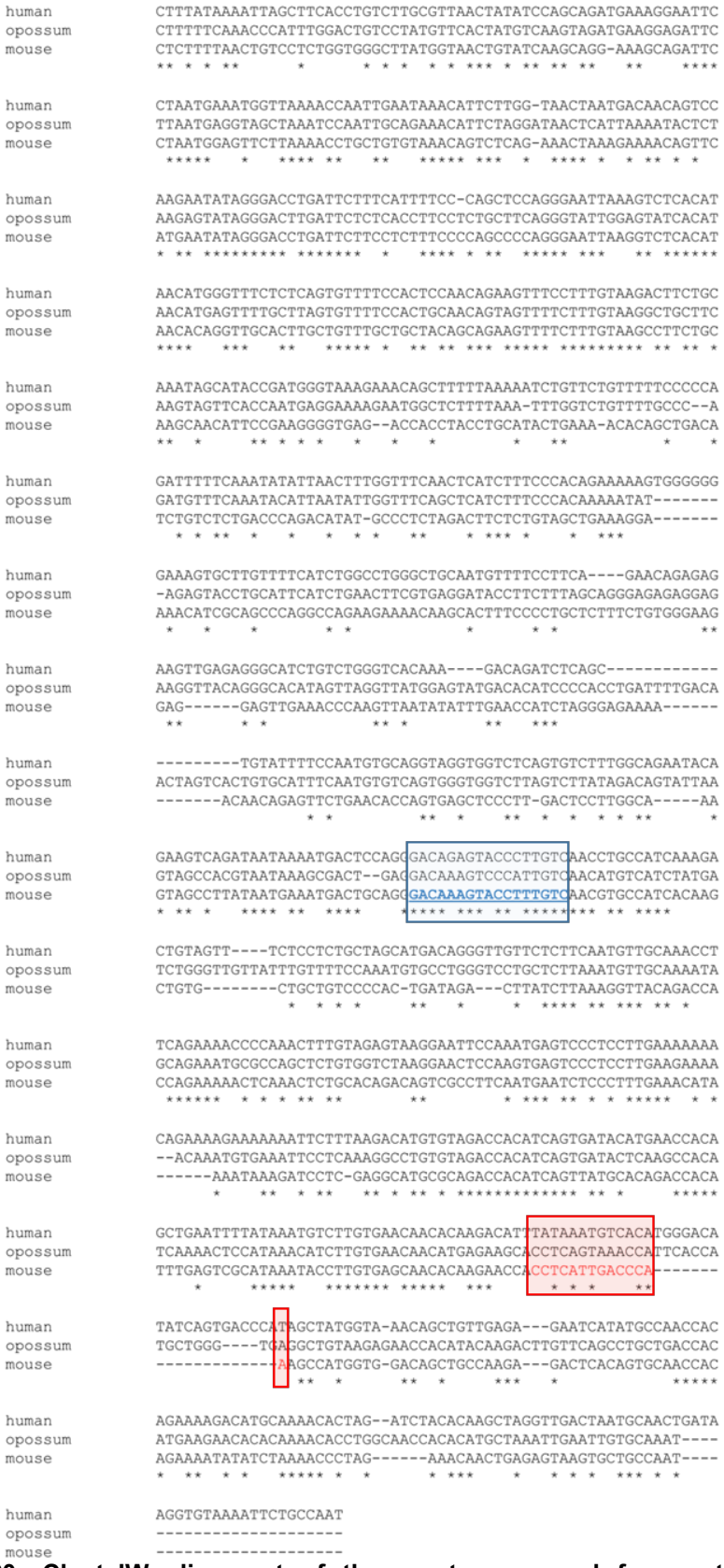


Figure 3.60: ClustalW alignment of the most conserved fragment of -230kb. Conservation of the sequence occurred to opossum. Bases that were identical between species when aligned are highlighted with asterisks. The TRAP predicted SOX9 motif (red box) was not evolutionarily conserved. The CAA and TTG bases of the putative SOX9 inverted repeat site (blue box) were conserved.

An oligonucleotide sequence spanning this inverted repeat and flanking sequence to 30bp was created for *in vitro* validation of binding interaction between DNA and sequence (Figure 3.61). A mutant oligonucleotide probe was also designed with disruption of the ACAA and TTGT motifs within the sequence.

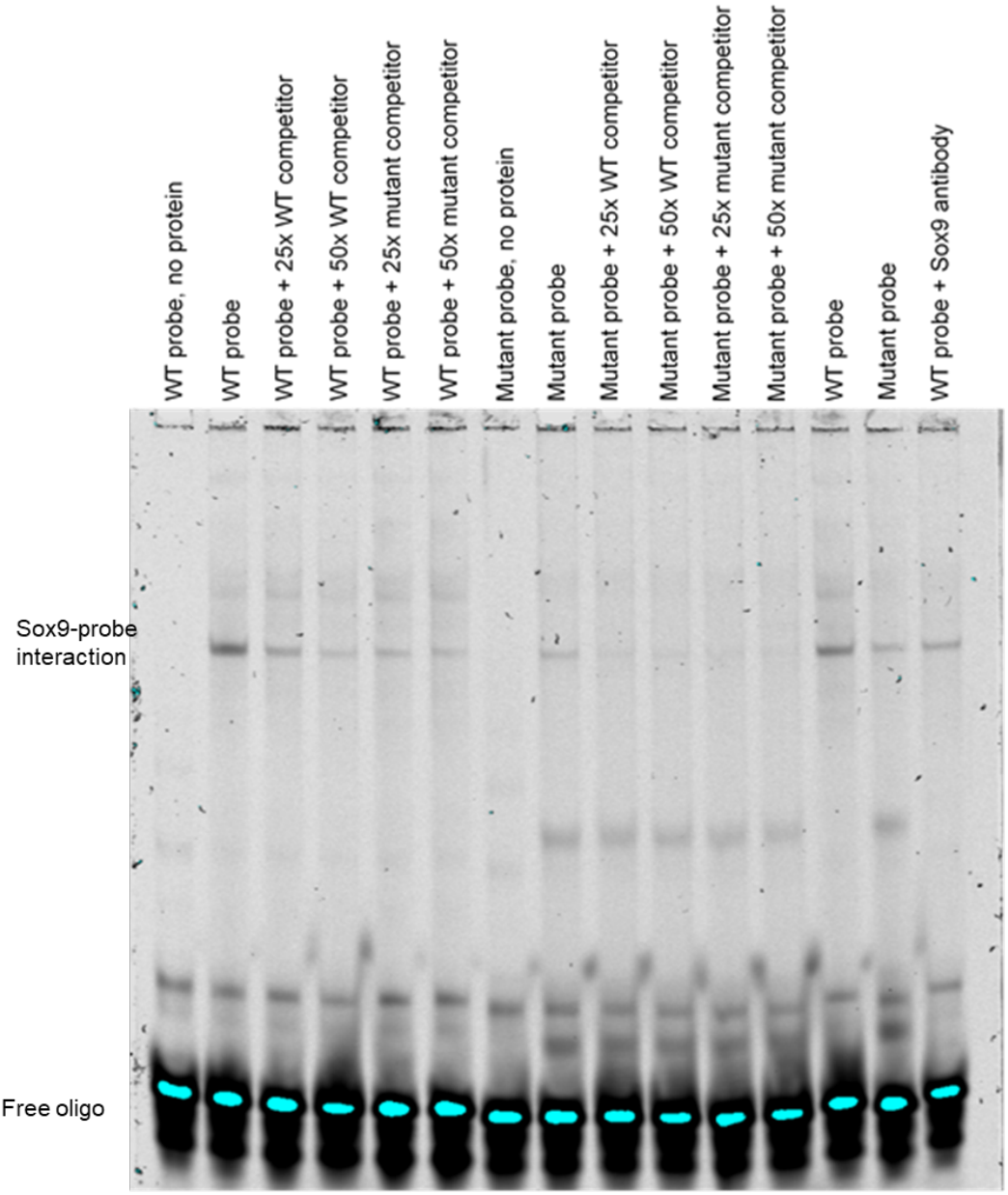


Figure 3.61: Determination of interaction between Sox consensus sequence within -230kb enhancer and SOX9 protein using EMSA. Bands represented shifted probe signal with SOX9 protein interaction were present in wells where reactions containing the wild-type probe were run. These bands were much fainter with the mutant probe, suggesting that interaction had not completely ablated DNA-protein interaction.

A band representing interaction between oligonucleotide sequence and SOX9 protein was discernible in reactions containing the wild-type probe, this band was much fainter in the mutant oligonucleotide containing reactions (Figure 3.61). The

band for the WT probe (lane 2, Figure 3.61) is more pronounced than those representing reactions that contained wild-type competitor. This suggests that the competitor prevented interaction between the probe and protein in a specific manner. The single band, rather than the doublet band that was observed for the -137kb and -148kb SOX9 suggests that SOX9 binds in monomeric form rather than dimeric form to interact with this -230kb TFBS (Oh *et al.* 2016).

3.8.4 RUNX2 does not bind strongly within -148kb

Two ChIP-Seq datasets pertaining to RUNX2 genomic interactions within MC3T3 pre-osteoblast and day 15 MC3T3 differentiated osteoblasts from Meyer (*et al*. 2014) (GSM1027478 and GSM1027496 respectively) were used in order to visualise potential interactions within enhancer regions. Peaks for these tracks within the UCSC Genome Browser lay within the -148kb enhancer region (Figure 3.62).

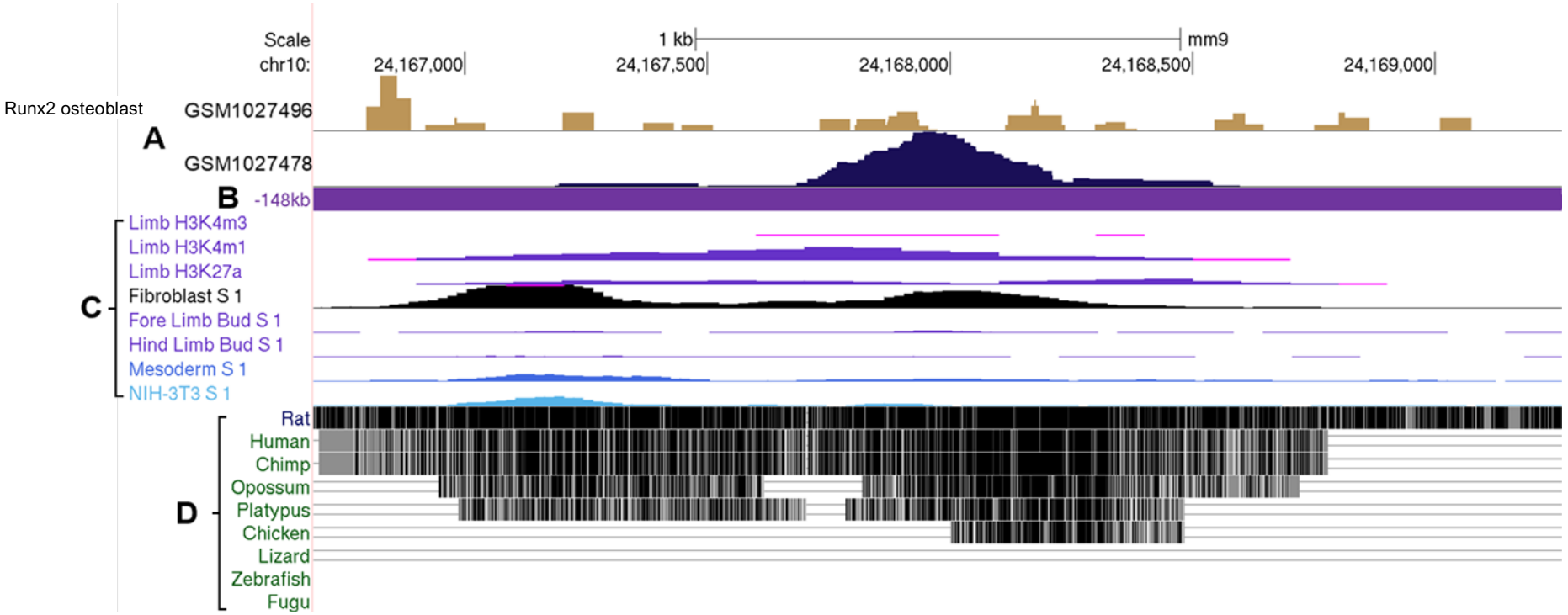


Figure 3.62: Visualisation of ChIP-Seq peaks for RUNX2 interaction within the -148kb enhancer. RUNX2 interacts within the enhancer in differentiated (GSM102796) and non-differentiated (GSM1027478) MC3T3 cells (Meyer *et al* 2014) (A). A large peak of interaction lay within a region of sequence conserved to chicken (D). Enhancer associated chromatin traits (C) were split into two clusters

Upon examination of the full -148kb enhancer sequence in TRAP (Chapter 5.5) a RUNX1 (AML) rather than RUNX2 TFBS was predicted within the RUNX2 ChIP-Seq peak for undifferentiated MC3T3 (Figure 3.62). There was no clear site for RUNX2 (OSF2) in TRAP predictions within this full sequence. RUNX2 binding sites were manually identified within the ChIP-Seq peak areas, with three TFBS probes for EMSA designed (Figures 3.63-65).

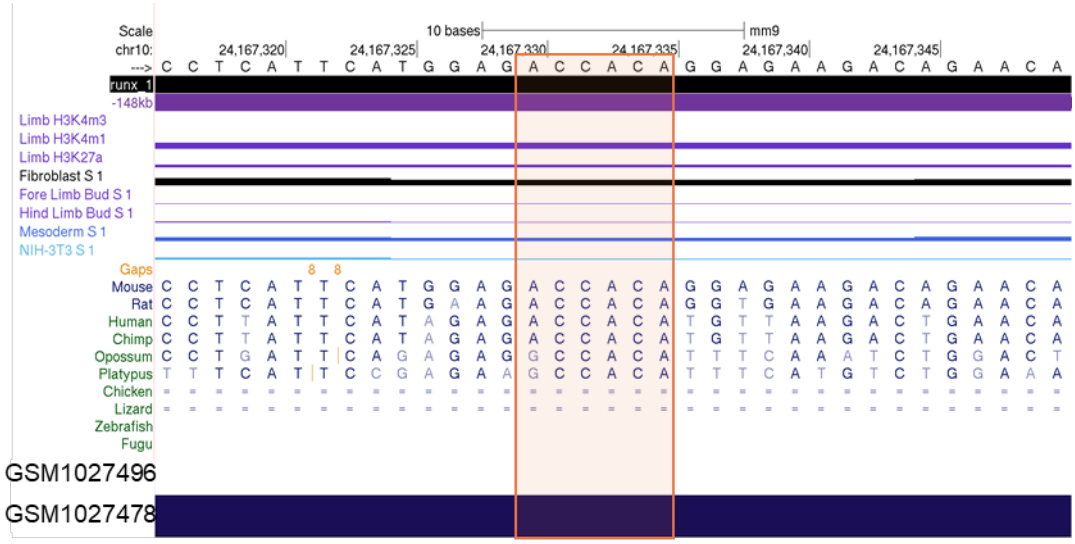


Figure 3.63: Visualisation of RUNX2_1 binding site within -148kb. RUNX2 consensus site ACCACA (orange box) was highly conserved between species. This formed the basis of an EMSA probe.

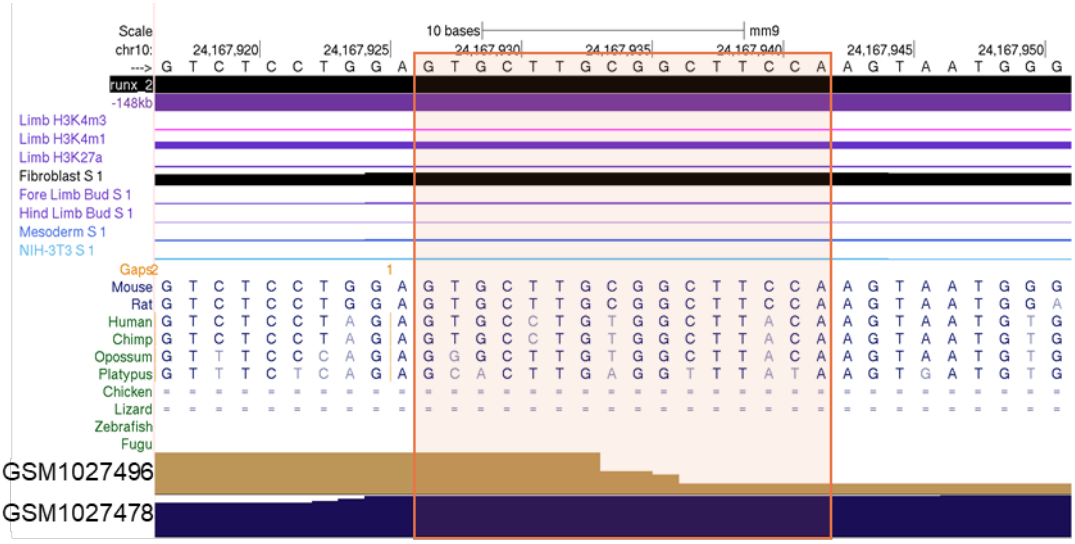


Figure 3.64: Visualisation of RUNX2_2 binding site within -148kb. Both of the ChIP-Seq datasets used peak within this sequence (Meyer *et al.* 2014) RUNX consensus site is highlighted with an orange box.

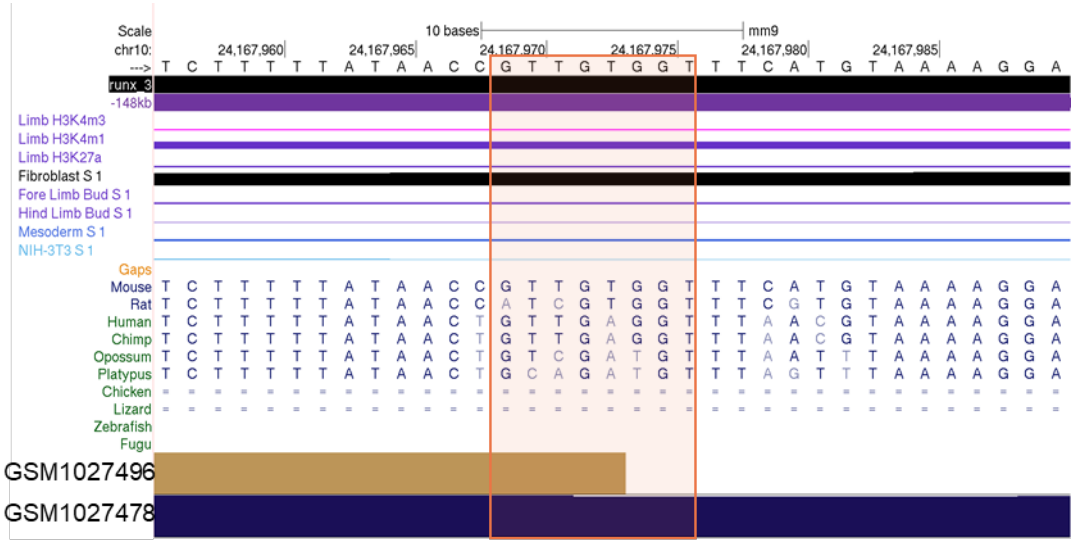


Figure 3.65: Visualisation of RUNX2_3 site within the -148kb enhancer. Both of the RUNX2 ChIP-Seq datasets used (Meyer *et al*. 2014) contain peaks within the probe in close proximity to conserved RUNX2 motif (highlighted with orange box).

The capacity for RUNX2 to bind each of these sequences was tested using EMSA. Binding conditions for DNA and protein were optimised using a control RUNX2 TFBS that has previously been validated in EMSA (Lamour *et al.* 2007) (Chapter 5.5) The band patterns for interaction between RUNX2 protein and probe for the RUNX probes designed herein did not match with the control but one of the bands for the RUNX2_3 probe was weaker with the use of a 25x and 50x excess of WT competitor, an effect that was reduced with use of a mutated competitor, suggesting specificity in interaction between probe and RUNX2 (Figure 3.66). *In silico* predictions of TFBS suggested the RUNX2_3 probe as most likely to be bound by Runx2. Unfortunately, a RUNX2 antibody was unavailable to conduct a supershift assay and further clarify DNA-protein interaction.

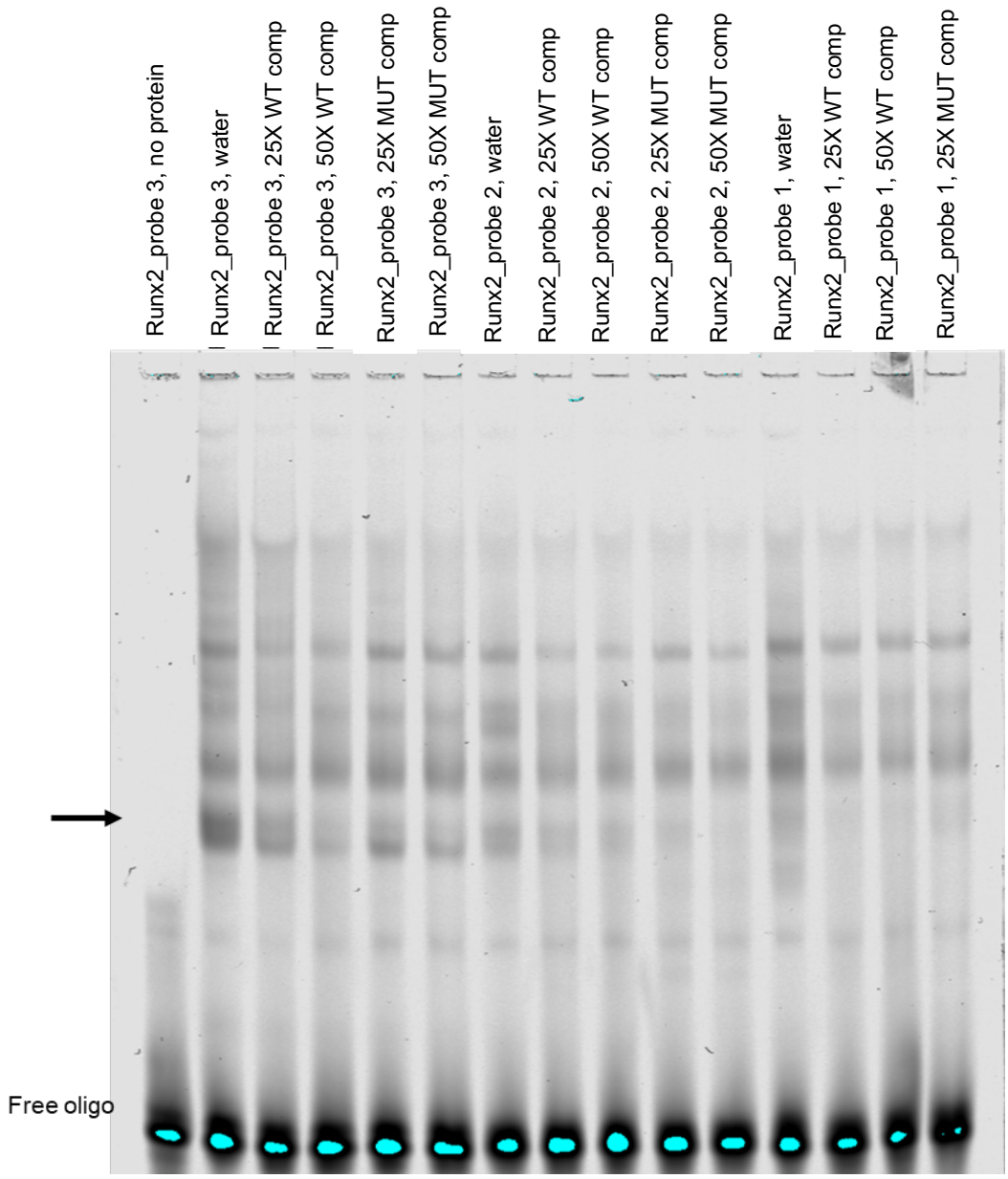


Figure 3.66: EMSA based investigation of interaction between RUNX2 consensus sequence within -148kb enhancer and RUNX2. An arrow highlights the band that was reduced for the RUNX2_3 probe with use of wild-type competitor, with both 25x and 50x excess, with reduced effect with the use of a mutant competitor.

3.9 CRISPR-Cas9 based manipulation of Ccn2 enhancer regions

As detailed in Chapter 2.4, CRISPR-Cas9 technology has revolutionised the capacity to manipulate genomic sequences, both *in vitro* and *in vivo.* In the current project, CRISPR-Cas9 technology was employed with the intention of excising enhancer regions from the murine genome *in vivo*, thereby generating transgenic mice with knockout of the *Ccn2* enhancers.

3.9.1 *In silico* identification and *in vitro* validation of CRISPR target sites

Guide RNA targeting loci in the vicinity of enhancer boundaries were identified using CHOP-CHOP tool (Labun *et al.* 2016; Montague *et al.* 2014); as outlined in Chapter 2, Materials and Methods 2.4.1. Guides were selected for the -137kb, -148kb and -230kb enhancers in addition to a region of interest spanning both -137kb and -148kb, as summarised in Table 3.10. (Guide sequence and off-targeting information is listed in appendices, Table 5.2).

Enhancer	Boundary	Guide names	Guide locus (mm10 genome)	Predicted efficiency	
				Xu *et al* (2015)	Moreno-Maetos *et al* (2015)
-137kb	5’	-137kb_5’_1	chr10:24457863-24457882	0.48	0.64
		-137kb_5’_2	chr10:24457862-24457881	0.42	0.64
	3’	-137kb_3’_1	chr10:24460119-24460138	0.60	0.61
		-137kb_3’_2	chr10:24460306-24460325	0.57	0.57
-148kb	5’	-148kb_5’_1	chr10:24447029-24447048	0.45	0.60
		-148kb_5’_2	chr10:24447022-24447041	0.44	0.60
	3’	-148kb_3’_1	chr10:24449585-24449604	0.48	0.59
		-148kb_3’_2	chr10:24448996-24449015	0.52	0.56
-148kb/-137kb	5’	-148kb/-137kb_5’	chr10:24446390-24446409	0.70	0.66
	3’	-148kb/-137kb_3’	chr10:24461139-24461158	0.53	0.59
-230kb	5’	-230kb_5’_1	chr10:24364823-24364842	0.71	063
		-230kb_5’_2	chr10:24364563-24364582	0.58	0.61
	3’	-230kb_3’_1	chr10:24366271-24366290	0.42	0.63
		-230kb_3’_2	chr10:24366337-24366356	0.38	0.70

Table 3.10: Identification of suitable gRNA for the deletion of enhancers. Two guides were chosen for the deletion of each single enhancer, with one each for 5’ and 3’ for the dual -148kb/-137kb deletion. The location of the target protospacer sequence and predicted efficiency scores in accordance with the work of (Moreno-Mateos *et al.* 2015; Xu *et al.* 2015)were recorded for each guide.

Where gRNA target sequences were identified within enhancer sequences; as oppose to flanking them, a rationale of compromise between gRNA efficiency and prediction of sequence function was made. This was based on the notion that sequences at the 5' or 3' extremities of the enhancer that are less conserved and would be unlikely to contain the lynchpin motifs that enhancer function depends on. Therefore, small fragments near the boundaries of enhancers left after CRISPR-Cas9 excision would not constitute enhancer function.

PCR products were designed to span target sites in order to assay cutting *in vitro.* Use of multiple primer pairs, including those used for the standard manipulation of enhancer sequences (Table 2.1) allowed assay of the gRNA to induce cutting in several substrates to be assessed. *In vitro* assay of enhancer cutting was carried out through incubation of target substrate PCR product, gRNA and recombinant SpCas9 before electrophoresis and visualisation of fragmentation. Initially, whole enhancer regions were used as substrate. This approach was later adapted with the use of smaller substrate for each enhancer boundary specifically rather than a whole enhancer region. Use of multiple primer pairs, including those used for the standard manipulation of enhancer sequences (Table 2.1) allowed assay of the gRNA to induce cutting in several substrates to be assessed. All combinations of primers used and resultant fragments generated through CRISPR-Cas9 genomic cleavage are listed in Appendices Chapter 5.3. The following results demonstrate that CRISPR-Cas9 cutting was more effective with the use of smaller substrates. This approach was therefore taken in the latter stages of the *in vitro* assays of CRISPR-Cas9 function. In addition, the concentration of reaction components was also optimised with the adoption of molar ratios of gRNA: Cas9 protein: substrate of 10:10:1 (as demonstrated in Appendices Chapter 5, Figure 5.18).

3.9.1.1 Excision of the -137kb enhancer

Two guide sequences were selected each for the 5' and 3' ends of the -137kb enhancer (Figure 3.65). The targets for the 5' boundary both resided within the enhancer sequence, however conservation in this sub region was low suggesting that the sequence was not important in the function of the enhancer. Therefore, if the 5' boundary and upstream sequence was retained upon CRISPR-Cas9 cutting, it would be highly unlikely that the enhancer function would be recapitulated. This ethos was also applied for 3' guide number 1 which sat approximately 140bp upstream of the 3' enhancer boundary.

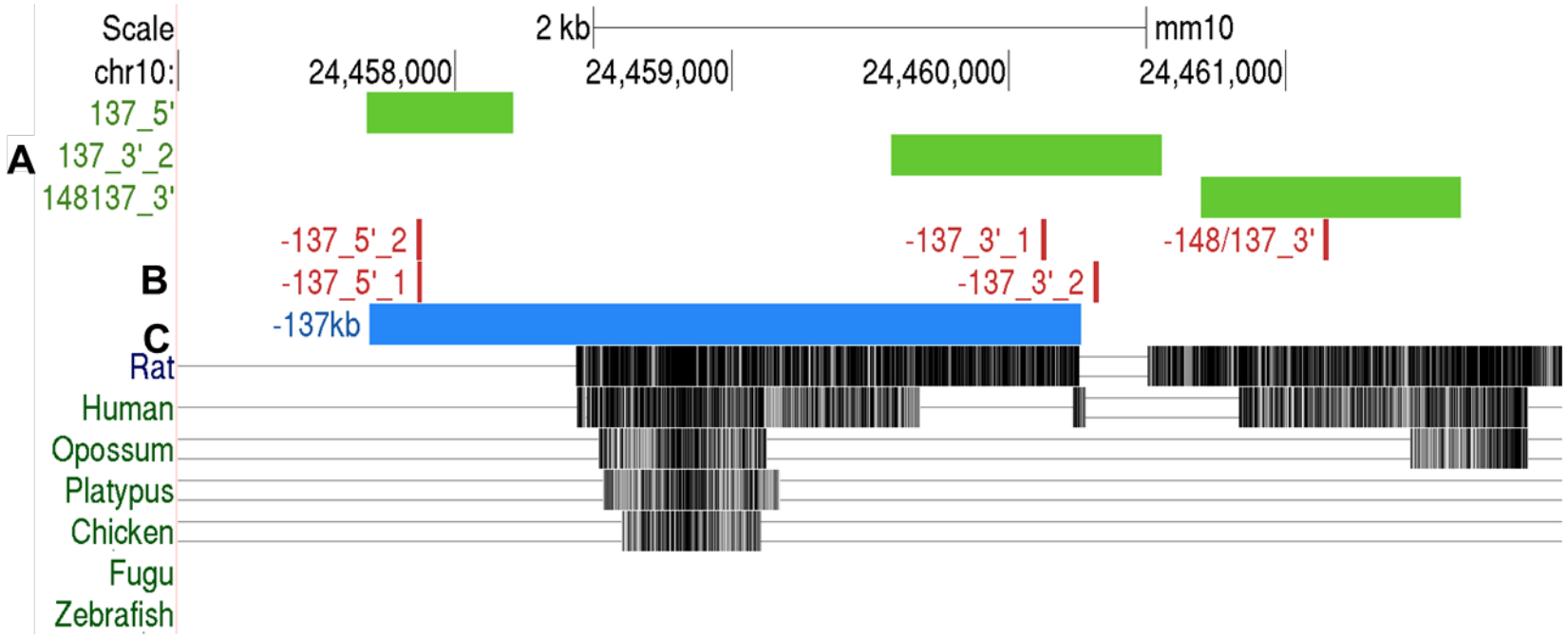


Figure 3.67: UCSC Genome Browser visualisation of CRISPR guide sequences targeting the -137kb enhancer. PCR amplicon were designed to span cut sites for assessment of cutting (green blocks, A)Two guides with the highest ranking for cutting efficiency (red bars, B) were selected for both the 5' and 3' ends of the enhancer region (blue block, C). Both of the 5' guide sequences sat approximately 200bp within the enhancer sequence. For the 3' end, 3'_1 sat approximately 130bp upstream of the enhancer boundary, whereas 3'_2 sat approximately 60bp downstream of the enhancer. The 3' guide for the -148kb/-137kb dual cut, -148/137_3' resides approximately 900bp downstream of the -137kb enhancer..

The capacity of each of these guides to induce DNA DSB in conjunction with the Cas9 was tested *in vitro.* This was firstly tested using a long substrate of 2864bp containing the full -137kb enhancer and additional 3' sequence in order to *incorporate 3' gRNA target site, as demonstrated in Figure 3.68.*

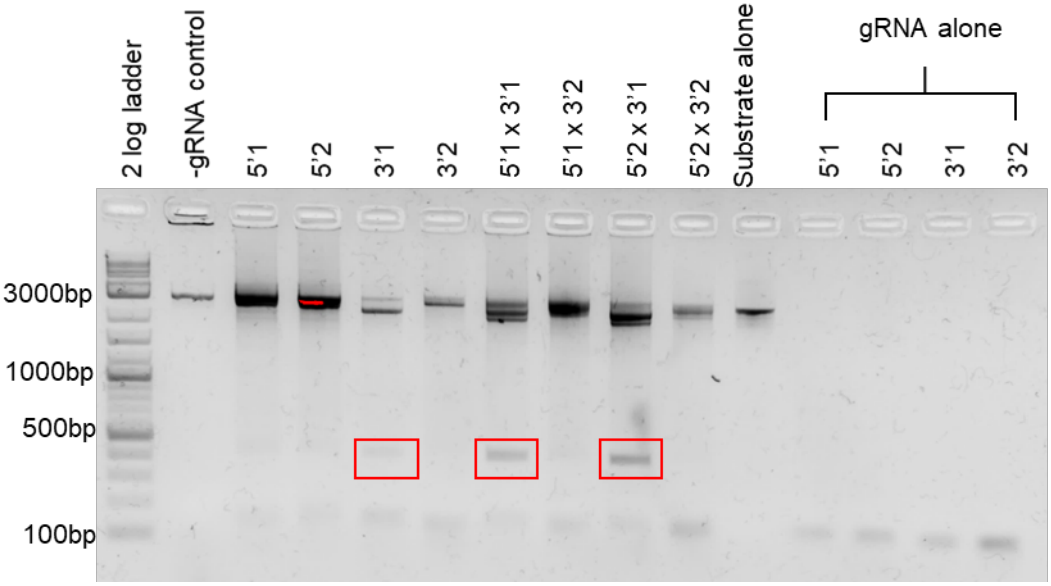


Figure 3.68: *In vitro* assay of -137kb gRNA function using full enhancer as substrate. Red boxes highlight bands corresponding to cleavage fragments. The faint bands between 100bp and 200bp represent gRNA, as demonstrated in the end 4 lanes which contained only gRNA. Substrate was cleaved through use of the -137kb 3'_1 guide RNA sequence only. Red marks in band indicate saturation with upper threshold of signal reached for transluminator system.

The -137kb_3'_1 guide facilitated cutting of the substrate with fragmentation of the PCR substrate corresponding to the loss of the 3' region. This occurred reproducibly in each reaction that contained this gRNA. Fragmentation in lanes corresponding to the activity of the other gRNA was not clear. An alternate approach with a shorter fragment was used to assess the capacity of the 5' gRNA to mediate Cas9 activity (Figure 3.69). The 529bp substrate was cut by both of the guides (red boxes, Figure 3.69), with gel electrophoresis fragmentation patterns tallying with DNA cut approximately 200bp and 330bp. The presence of uncut substrate in both of the lanes containing gRNA indicates that cutting was not 100% efficient, although the band corresponding to uncut substrate was fainter than in the –gRNA control lane. The intensity of cut fragments was similar for both gRNA suggesting that gRNA-Cas9 interaction and cutting efficiency was similar for each gRNA.

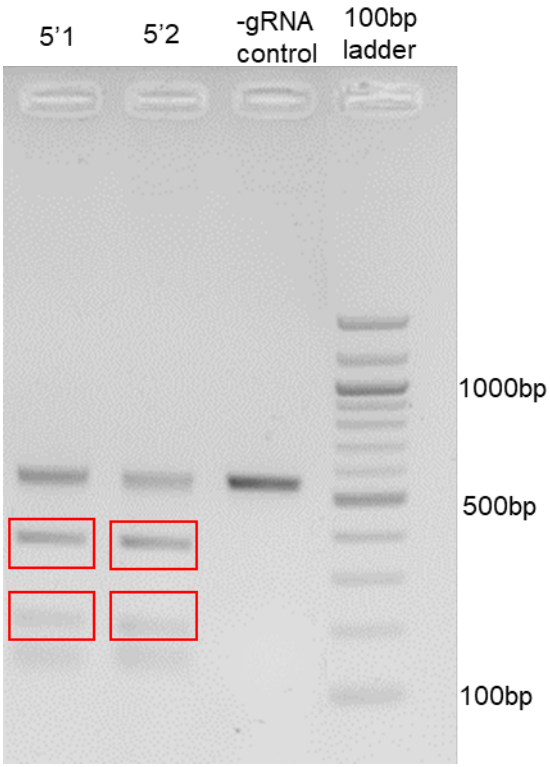


Figure 3.69: *In vitro* assay of -137kb 5' guides capacity to cut PCR amplicon spanning 5' enhancer boundary. Red boxes highlight bands corresponding to fragments produced from successful cleavage of the 529bp substrate. Both guides facilitated cleavage of the substrate DNA at the target site. Faint bands are also present at approximately 140bp corresponding to the guide RNA.

3.9.1.2 Excision of the –148kb enhancer

For the 148kb 5' region, both guides sat within the enhancer sequence, approximately 160bp downstream of the boundary. For the 3' guides, guide 1 sat downstream of the 3' boundary, whereas guide 2 sat approximately 400bp upstream of the 3' boundary in sequence with little interspecies conservation. The location of each guide with respect to the enhancer is illustrated in Figure 3.70.

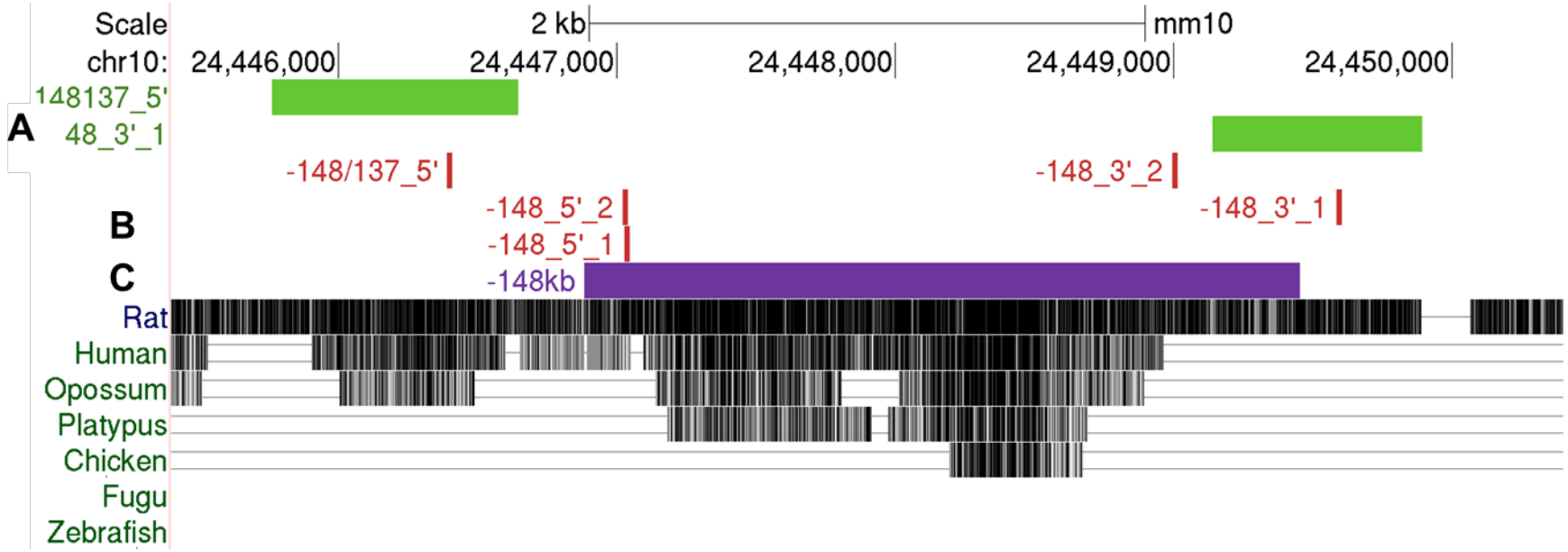


Figure 3.70: UCSC Genome Browser visualisation of CRISPR guide sequences targeting the -148kb enhancer. Both of the 5' guide sequence (red bars, B) lie within the enhancer (purple bar, C), approximately 160bp downstream of the 5' boundary. For the 3' guides, 3'_2 resides approximately 450bp upstream of the 3' boundary within the boundary sequence whereas the 3'_1 guide sits approximately 150bp downstream of the enhancer. For each guide target within the enhancer, the sequence is not highly conserved between species, suggesting these regions are dispensable in enhancer function. The -148kb/-137kb dual cut 5' guide is found approximately 500bp upstream of the enhancer region. PCR amplicon were designed for the -148kb/-137kb 5' and -148kb_3'_1 guides (green bars, A). For the other guides, substrate of the whole enhancer was used for in vitro validation of CRISPR-Cas9 activity.

For the assay of gRNA activity, substrate of the basic enhancer region was carried out first (Figure 3.70). Albeit faint on the scanned gel image (rather than from digital image file), the fragment in the -148kb_5'_1 lane corresponds to the 5' enhancer boundary to gRNA target site of approximately 160bp. This suggests that the gRNA is functional, but the efficiency of cutting is low. This fragmentation was repeated with the dual assay of -148_5'_1 and -148kb_3'_2, with fragmentation pattern showing that the 5'_1 guide alone facilitated substrate cleavage.

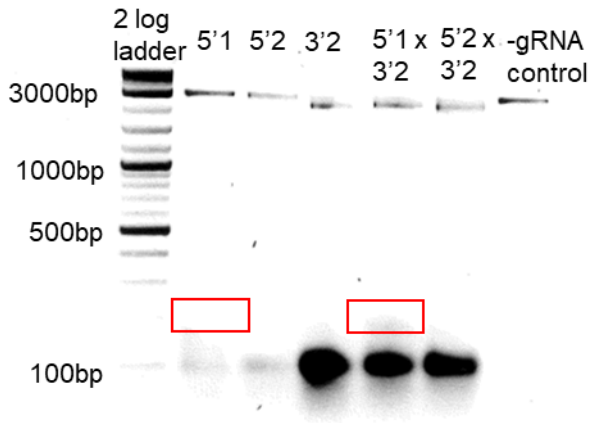


Figure 3.71: Initial assay of -148kb 5'_1, 5'_2 and 3'_2 guides capacity to facilitate Cas9 cutting of substrate DNA. Bands representing cleavage product of the substrate were visible (within the red boxes) with the use of the -148kb 5'1 gRNA. The most substantial bands however represent guide RNA at approximately 120bp.

This experiment was repeated with the use of the -148kb_3'_2 gRNA (Figure 3.72). Two substrates were used, firstly the standard -148kb enhancer region for 5'_1, 5'2, 5'_1x 3'_2 and 5'_2x 3'_2 reactions, in addition to a substrate spanning from the 5' enhancer boundary to 3' of the 3' cutting substrate reverse primer.

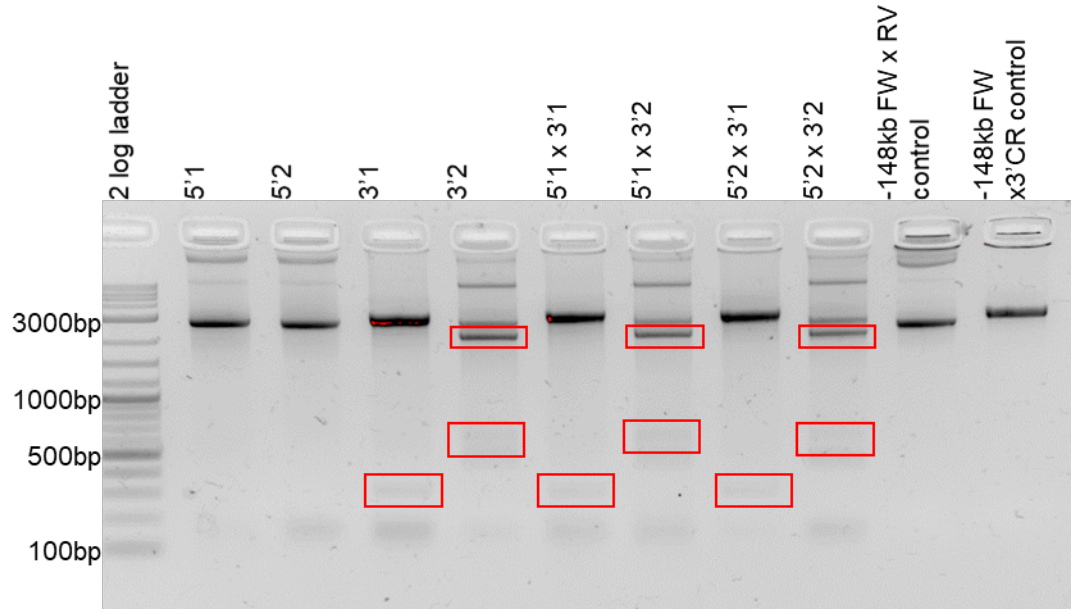


Figure 3.72: *In vitro* assay of capacity of -148kb guides to facilitate cleavage of -148kb enhancer containing substrate. Amplicon of the basic enhancer region was for each reaction aside from those containing -148kb_3'1 which used the enhancer forward primer and reverse primer for 3'C substrate. Use of both 3' guides led to cleavage of substrate DNA with correct fragment sizes (highlighted with red boxes).

Given that neither 5' guide had generated discernible fragments in this experiment a further substrate was used spanning further upstream of the 5' enhancer boundary (-148kb/-137kb 5' cutting forward primer x reverse primer in the middle of the -148kb region). Both guides led to the cutting of the substrate, with fragmentation tallying with expected split of substrate into fragments of approximately 1270bp and 1700bp (Figure 3.73). Using this approach fragmentation of substrate corresponding to the action of each gRNA was produced.

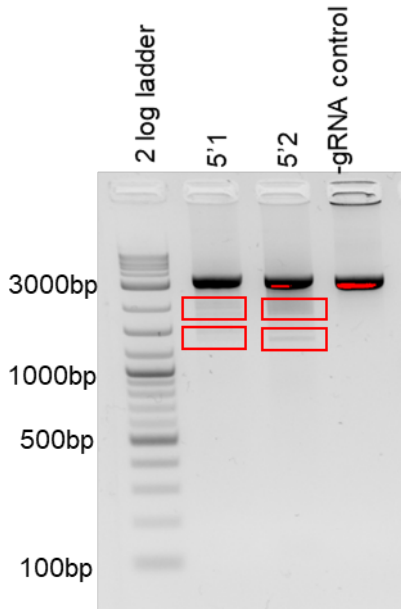


Figure 3.73: *In vitro* assay of -148kb 5' gRNA mediated cutting of substrate containing the -148kb enhancer. Substrate was amplified using the -148kb/-137kb 5' cutting assay forward primer and reverse primer recognising sequence within the enhancer.

3.9.1.3 Excision of the –230kb enhancer

For the 5' guides, _1 sat approximately 180bp within the enhancer sequence, whereas _2 was localised approximately 90bp upstream of the 5' enhancer boundary. For the 3' guides, both sequences sat at least 140bp downstream of the enhancer (Figure 3.74).

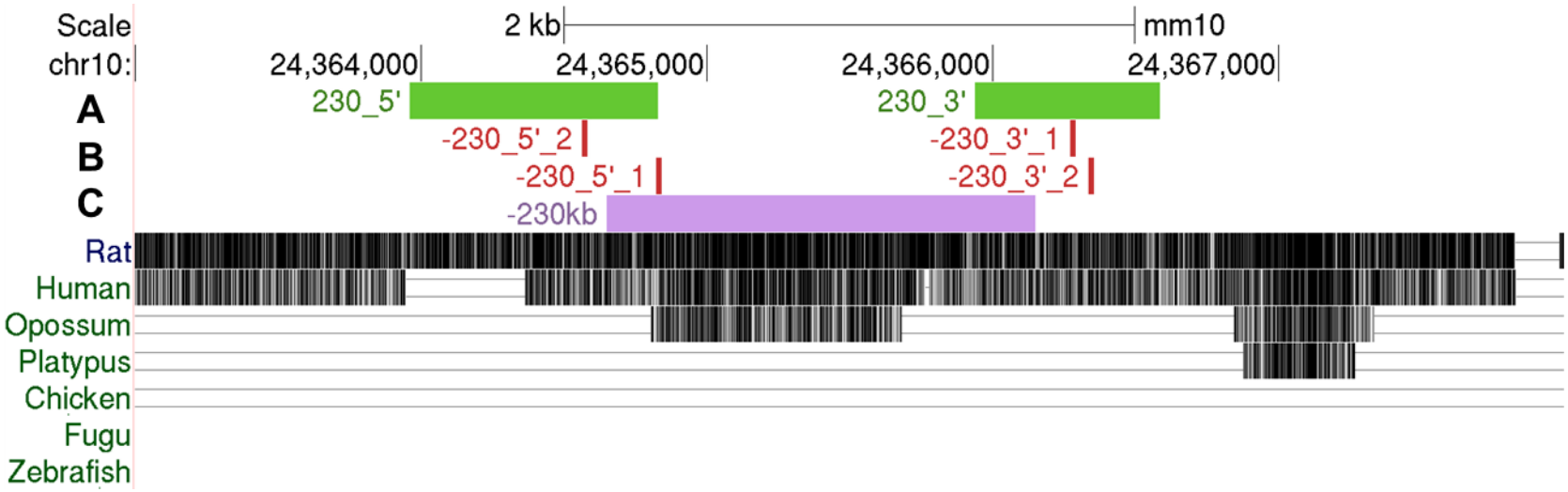


Figure 3.74: UCSC Genome Browser visualisation of CRISPR guide sequences within the -230kb enhancer. Aside from -230kb_5'_1, each target sequence gRNA (red bar, B) was found within sequence flanking the -230kb enhancer (lilac block, C) region. The target site for 5'_1 lay approximately 180bp downstream of the enhancer 5' boundary, with a compromise made between predicted efficiency and sequence conservation. Primers were designed for *in vitro* validation, with amplicon spanning the -230kb_5'_2 target region and both 3' target loci (green bars, A).

For the -230kb enhancer, multiple experiments were carried out in order to validate gRNA function. Initially, a 1950bp substrate spanning from the 5' enhancer boundary to the 3' end of the 3' cutting substrate amplicon was used to test the 5'_1 guide whereas substrate of the standard -230kb region was used to test the 5'_2 and 3'_1 guides individually in addition to this gRNA in combination with one another (Figure 3.74).

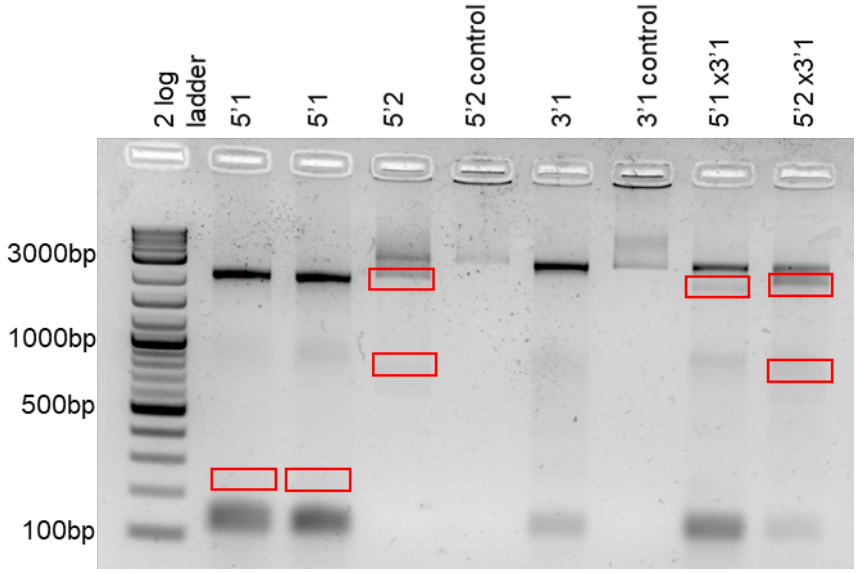


Figure 3.75: Initial assay of ability of -230kb 5'_1, 5'_2 and 3'_1 gRNA to induce cutting of DNA *in vitro*. The use of the 5'_1 and 5'_2 led to cleavage of substrate with fragments (denoted by red boxes) corresponding to target site DSB. The 3'_1 guide did not mediate cutting of the substrate.

The capacity of the 5'2 guides to facilitate DNA cutting was further tested using a substrates designed to span each enhancer boundary. There was digestion of an amplicon spanning 869bp around the 5'_2 guide (Figure 3.76).

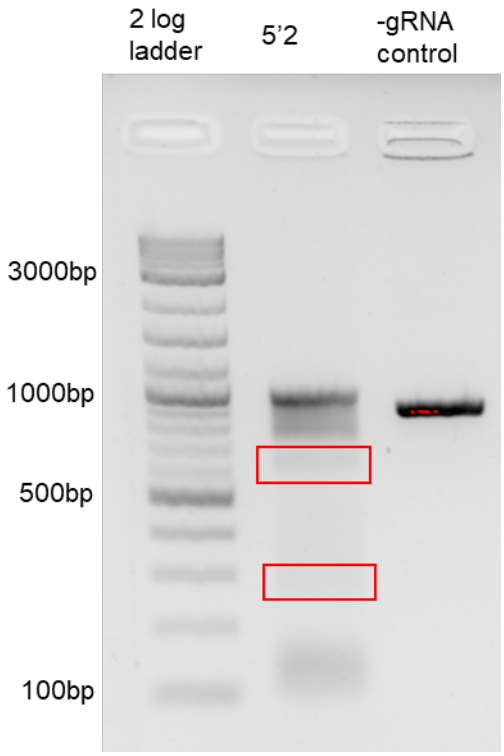


Figure 3.76: *In vitro* digestion of -230kb 5' cutting assay substrate with -230kb_5'_2 guide. PCR amplicon substrate was designed to span the -230kb_5'_2 target span. The fragment bands produced (asterisk) correspond to correct cleavage by Cas9 at the target site. The faint, smeared band above 100bp corresponds to gRNA.

This process of using smaller amplicon designed to span the gRNA targeting site was repeated for the -230kb_3' guides (Figure 3.77). The -230kb_3'_1 guide functioned *in vitro* (red boxes Figure 3.77), whereas the -230kb_3'_2 guide did not. Moreover, the band for uncut 648bp substrate on the gel was much fainter than that in the –gRNA control, suggesting that Cas9-gRNA interaction functioned with high efficiency.

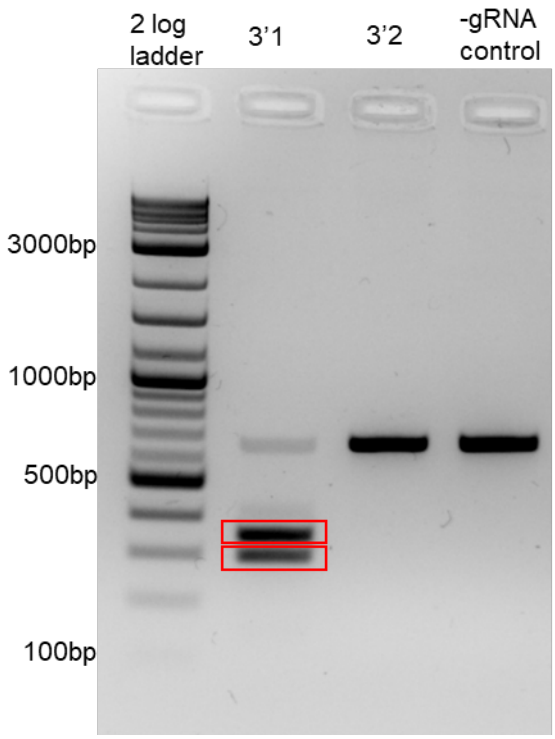


Figure 3.77: *In vitro* assay of -230kb 3' guide function using substrate spanning the gRNA target site. The -230kb_3'_1 gRNA efficiently mediated cleavage of the substrate by Cas9, with pronounced bands corresponding to fragments generated by DSB (red boxes). Moreover, the band representing uncut substrate within this well was faint. The -230kb_3'_2 guide did not facilitate DNA cleavage, with no fragmentation visible in the gel lane and band representing uncut substrate as in the –gRNA control.

3.9.2 CRISPR mediated removal of enhancers *in vivo*

Microinjection of mouse zygotes (as detailed in Chapter 2.5) was carried out in order to generate transgenic mice in which enhancer regions had been excised from the genome. Validated gRNA and recombinant Cas9 were injected into the cytoplasm of one-cell stage embryos (0.5dpc). The number of embryos harvested and injected, in addition to survival rate is recorded in Table 3.10. The number of mice born, and that survived to genotyping in addition to genotyping results are also recorded in Table 3.11. None of the mice born exhibited any discernible phenotype that could be associated with genomic manipulation. Three lines; Ccn2Δ148/137, Ccn2Δ148 and Ccn2Δ230 were generated.

Line (enhancer)	Guide RNA used	Number of embryos isolated	Number of embryos injected	Number of embryos survived	Number of pups born	Number of mice genotyped	Number of mice with deleted copy of enhancer
Ccn2Δ148 (-148kb)	-148kb_5'_1 and -148kb_3'_1	220	200	130	28	27	0
Ccn2 Δ148/137 (-148kb/ -137kb dual)	-148kb_5'_1, -148/137_5, -137kb_3'_1 and -148/-137_3'	270	250	150	50	46	0
Ccn2Δ230 (-230kb)	-230kb_5'_2 and -230kb_3'_1	100	80	56	28	27	0

Table 3.11: Record of microinjection of gRNA and Cas9 into mouse embryos, and subsequent survival of mice to genotyping stage, and number of mice genotyped as exhibiting deletion of the enhancer sequence.

Mice were genotyped through PCR of amplicon within the enhancer regions, in addition to spanning gRNA targeting sites (Chapter 5, Tables 5.4 to 5.7) Using this approach, successful CRISPR-Cas9 removal of the enhancer region would prevent amplification of region of interest in PCR. Product formation was assessed on agarose gels, with an internal control for every sample in order to rule out false positives for enhancer knockout. In the following gel images, where there is no band for internal control, PCR was repeated with lower concentration of DNA, DNA was re-precipitated or mouse was re-notched for sample. Therefore, there was an internal control reaction for every sample, albeit not on the same gel as amplicon of interest product. PCR conditions for each amplicon were optimised using genomic DNA isolated in the same manner as genotyping. However, multiple amplicon had to be tested for each line in order to ensure result accuracy.

3.9.2.1 Genotyping of Ccn2Δ148/137

The accuracy of genotyping PCR was a major issue in the genotyping for the simultaneous knockout of the -148kb and -137kb enhancers in the Ccn2Δ148/137 line. A schematic of the PCR strategy used is illustrated in Figure 3.78.

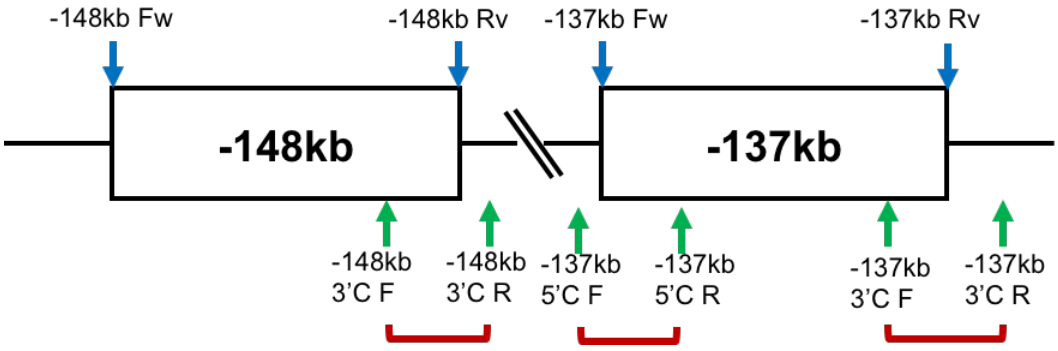


Figure 3.78: schematic of genotyping strategy for Ccn2Δ148/137. Standard cloning PCR primers are indicated for context with blue arrows. However, the CRISPR assay substrate primer pairs (green arrows) were used in attempts to generate amplicon (red bars) and therefore confirm whether the enhancers had been deleted.

The problems that were faced are typified by the results of genotyping animals 9.1 to 13.2 of the line (Figure 3.79). Both the -137kb 5' and -148kb 3' CRISPR cutting assay amplicon were used with optimised conditions. The -148kb 5' amplicon was not produced in any reaction which indicated a failure in PCR rather than 100% efficiency of the CRISPR method; which is highly unlikely. For some of the samples, the -137kb 5' product was absent, indicating a knockout. Where bands -137kb 5' CRISPR cutting assay were faint, but the internal control band was comparable to

others, it would suggest heterozygosity. If the CRISPR-Cas9 system removed the sequence from one chromosome, a single copy of the target would still be present in the genome and would be amplified by PCR; albeit to a lesser extent than in a homozygote wild-type sample.

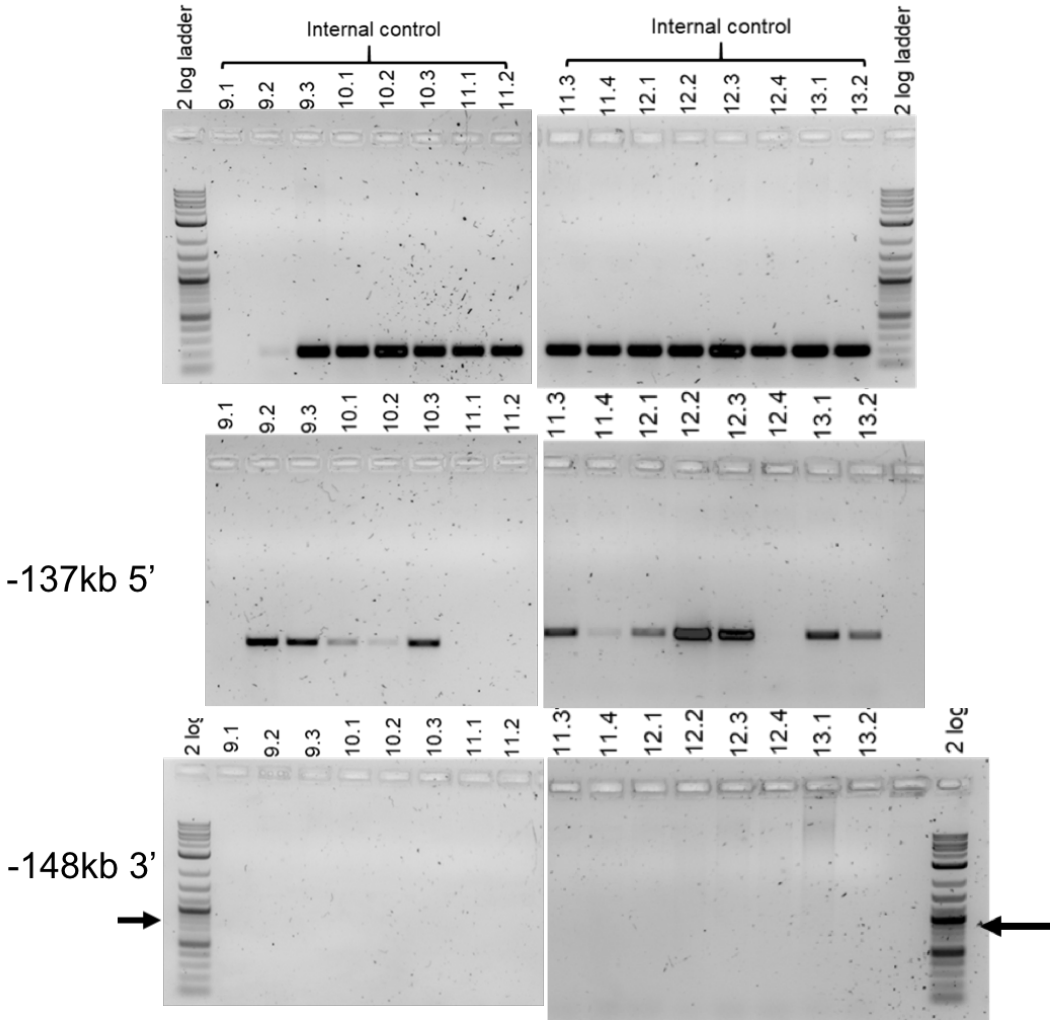


Figure 3.79: Example of genotyping from Ccn2Δ148/137 line. Internal PCRs were carried out for every sample in order to ratify template suitability and reduce the possibility of false positives. Black arrows denote the expected band size of PCR amplicon. These reactions were repeated for samples 9.1 and 9.2 in order to ensure DNA integrity. The -148kb 3'C amplicon was not produced in any sample suggesting issue with PCR experimental set up rather than CRISPR deletion.

These samples were used as template in further PCR reactions with alternate amplicon. In these reactions, samples that were previously genotyped as potentially being heterozygous or homozygous for enhancer knockout were genotyped as still containing the region of interest (Figure 3.80).

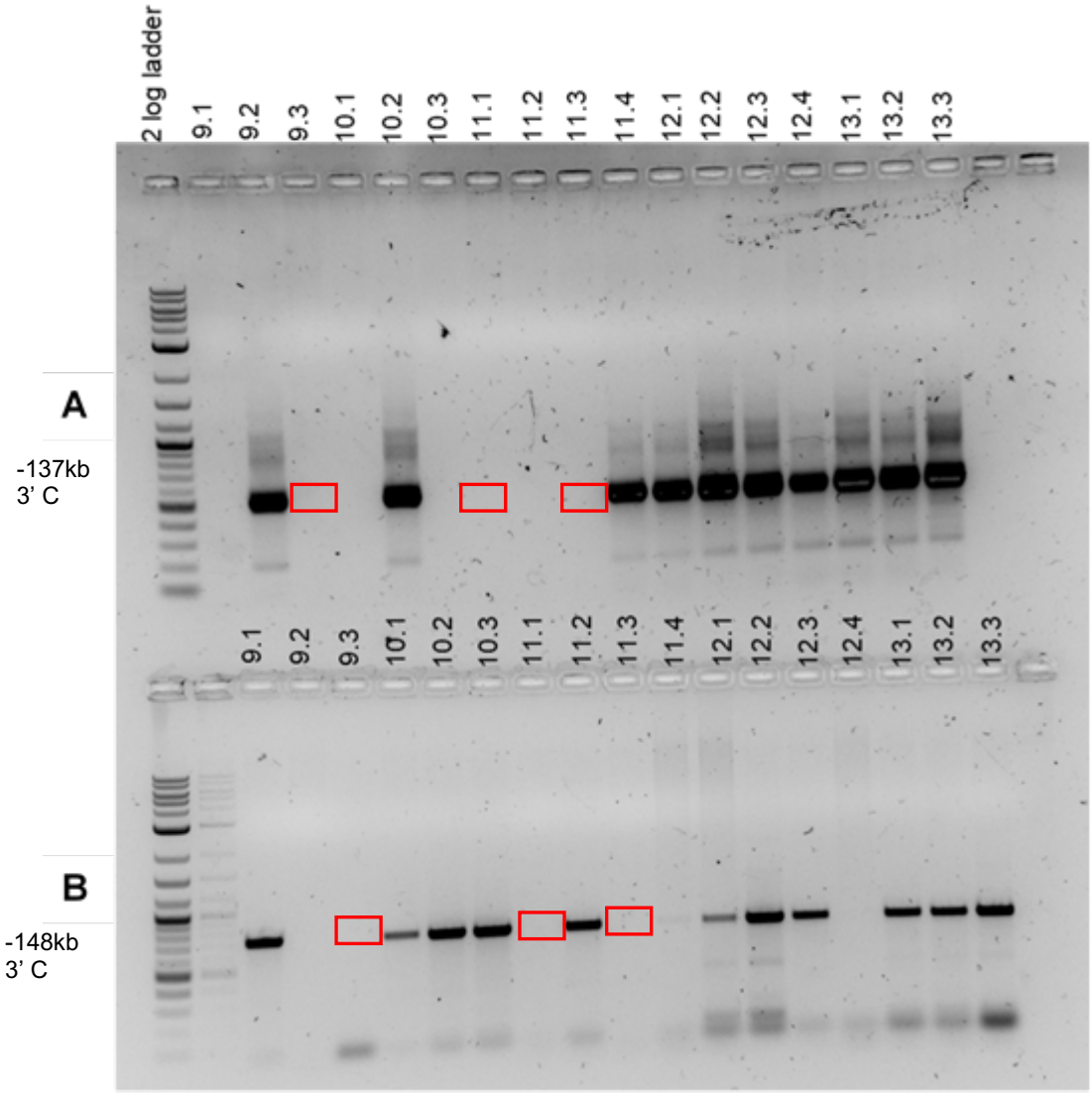


Figure 3.80: Further genotyping of mice from the Ccn2Δ148/137 line. The -137kb 3' CRISPR cutting substrate (529bp) was used as amplicon in gel half A, whereas the CRISPR substrate primers for -148kb 3' (752bp) were used for gel half B. There was no product from sample 9.3 or 11.1 and 11.3 (red boxes) for either amplicon, suggesting CRISPR may have been successful.

A total of eight different PCR conditions were tried in order to ascertain whether any mouse was negative for each amplicon. 9.1, 11.3 and 12.4 were all chosen as the most feasible heterozygotes for the dual deletion of -148kb/-137kb. Unfortunately, these mice were all male and so propagation of the line would require crossing with homozygous wild-type females. This was not a feasible option as resultant offspring would be heterozygous at best which could not be efficiently interrogated in the scope of this project in terms of time and budget and so propagation of other transgenic lines was prioritised.

3.9.2.2 Genotyping of Ccn2Δ148 transgenic line

As with the Ccn2Δ148/137 line, problems with PCR hampered the accuracy of Ccn2Δ148 line genotyping. A schematic of the genotyping approach used for this transgenic mouse line is outlined in Figure 3.81

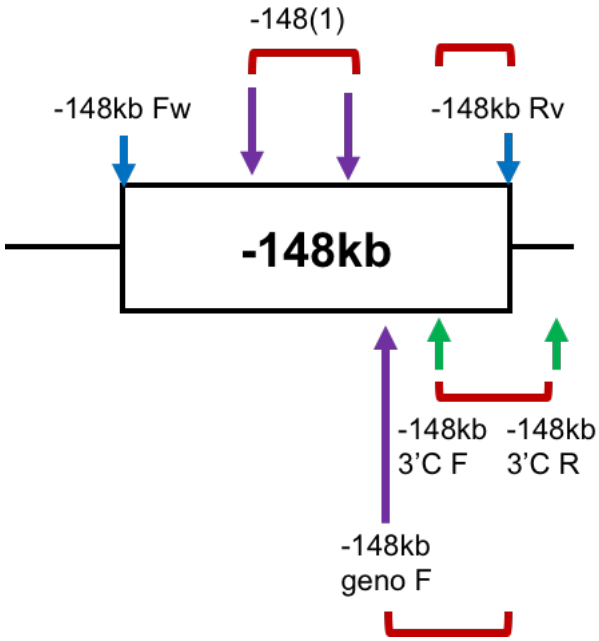


Figure 3.81: Schematic representation of the genotyping of the excision of the -148kb enhancer in Ccn2Δ148 transgenic line. The positions of primers used for cloning of the region are highlighted with blue arrows. The green arrows indicate position of primers for CRISPR *in vitro* assay substrate amplicon and the purple arrows indicate primers that amplify from within the enhancer that were used for genotyping of the *Lac*Z reporter transgenic line. Amplicon generated for genotyping are highlighted with red bars.

Initial genotyping is illustrated in Figure 3.82.

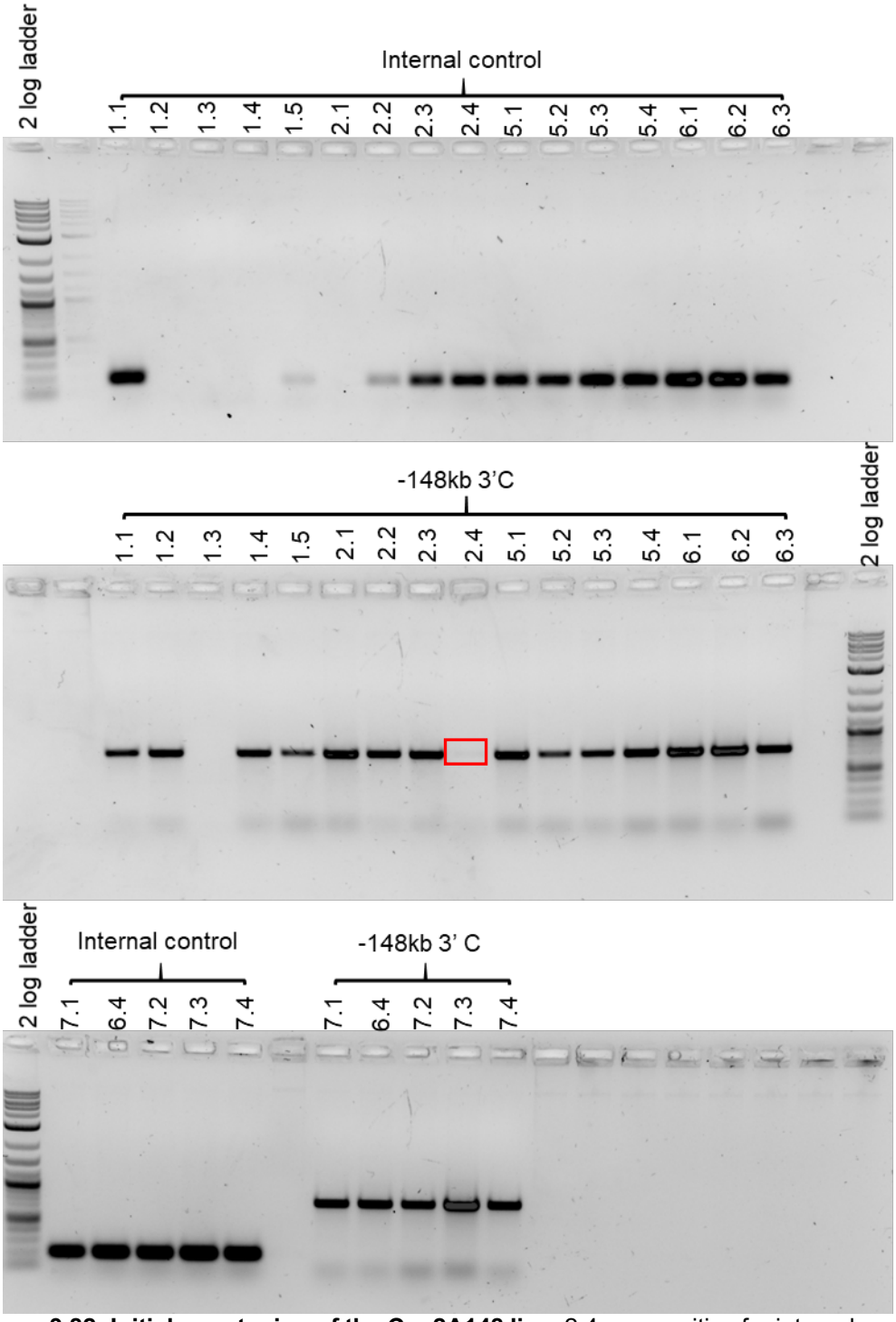


Figure 3.82: Initial genotyping of the Ccn2Δ148 line. 2.4 was positive for internal control reaction, but negative for the amplicon of interest (red box), suggesting that this mouse harboured knockout of the enhancer.

Genotyping PCR was repeated with several amplicon of interest where results were unclear from initial genotyping (as shown in Figure 3.83).

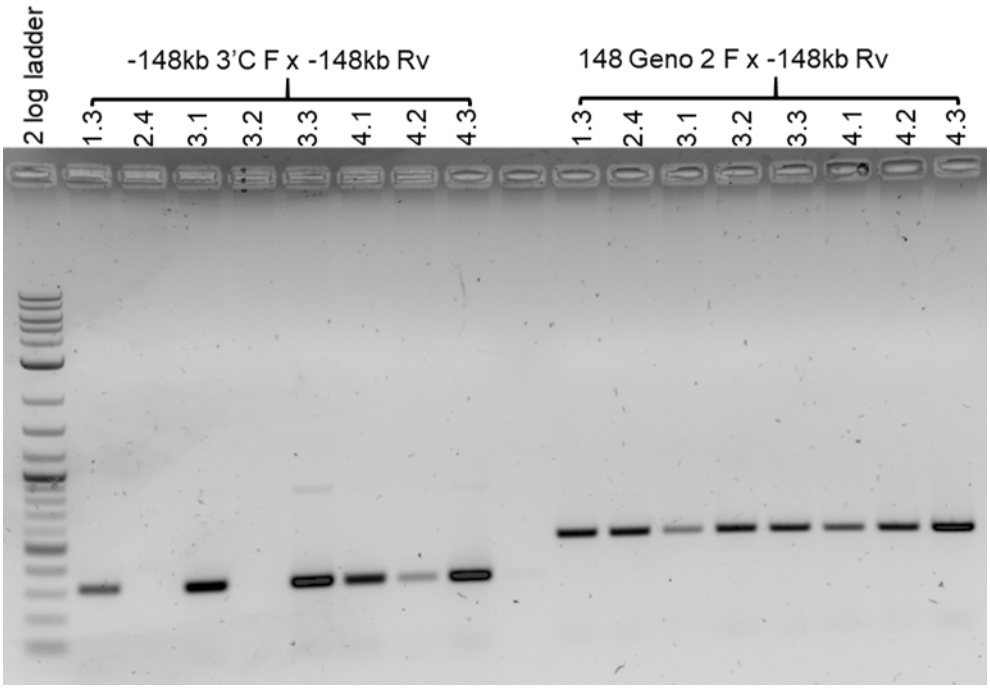


Figure 3.83: Representative gel image further genotyping of the Ccn2Δ148 line. Sample 1.3 which was previously negative for enhancer amplicon (Figure 3.79) contained both amplicon sequences. Bands that were faint and therefore suggesting heterozygosity for enhancer deletion were not concordant between PCR for samples, suggesting this difference was due to PCR conditions.

Across a total of six genotyping PCR experiments, 2.4 and 3.2 were chosen as potential heterozygotes and bred to propagate the line, with the hope of producing homozygote knockouts. F_1 mice were genotyped, as demonstrated in Figure 3.84.

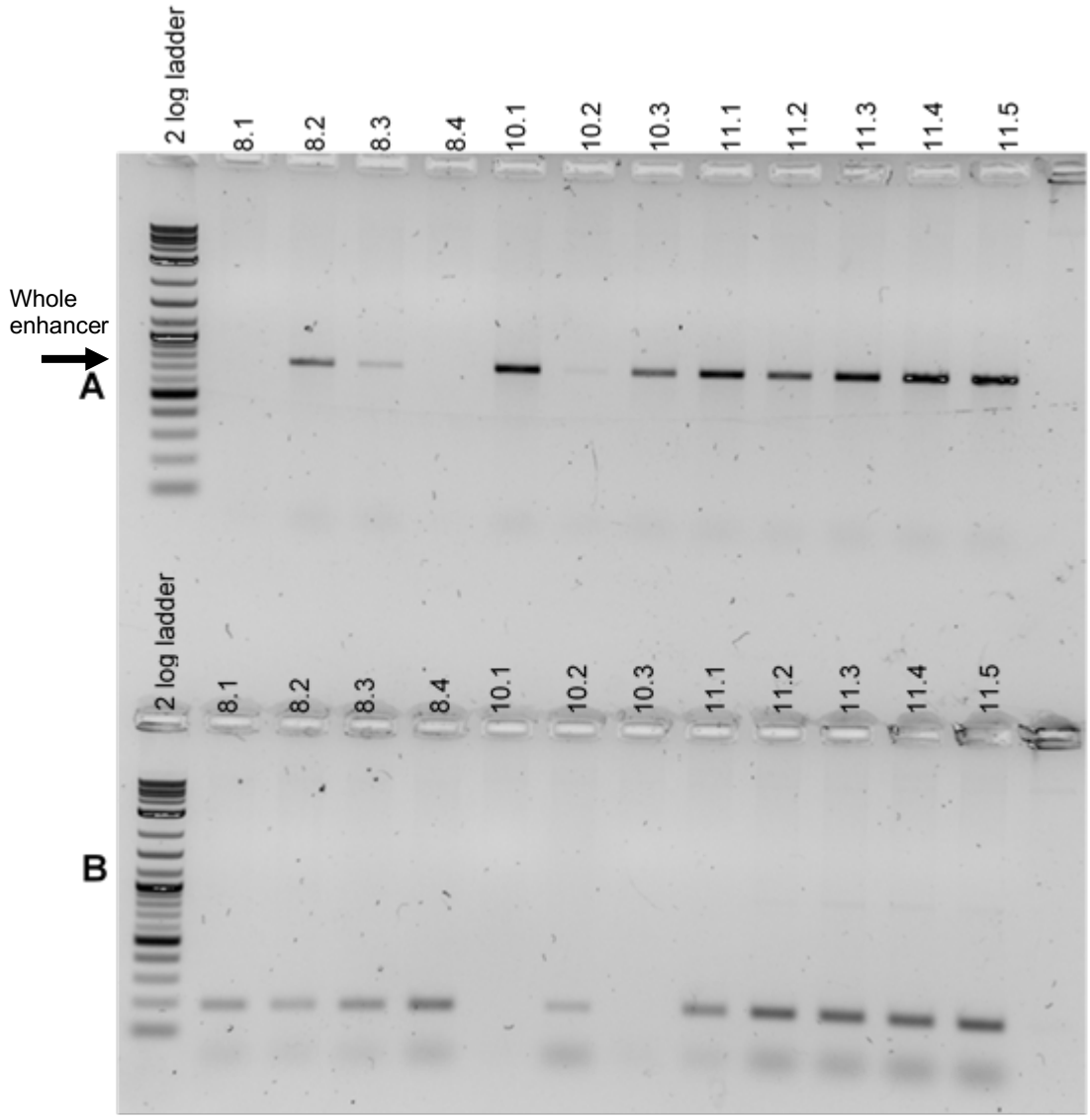


Figure 3.84: Genotyping of F1 from Ccn2Δ148 line. Mice were genotyped for an amplicon spanning the middle of the enhancer (148-1) (A) in addition to internal control (B). There was a problem with the 10.2 template in the 148-1 reaction which is therefore a false positive for knockout. None of the mice were homozygous for deletion of the enhancer as shown by product formation for the 148-1 reaction.

These mice were also genotyped from tissue after culling, with all samples producing PCR amplicon and therefore containing the -148kb enhancer. At this point, further propagation of the line was not temporally or fiscally feasible and so the line was discontinued.

3.9.2.3 Genotyping of Ccn2Δ230 transgenic line

The genotyping of Ccn2Δ230 mice was also problematic. A schematic of the approach used to PCR for amplicon within the enhancer region is shown in Figure 3.85.

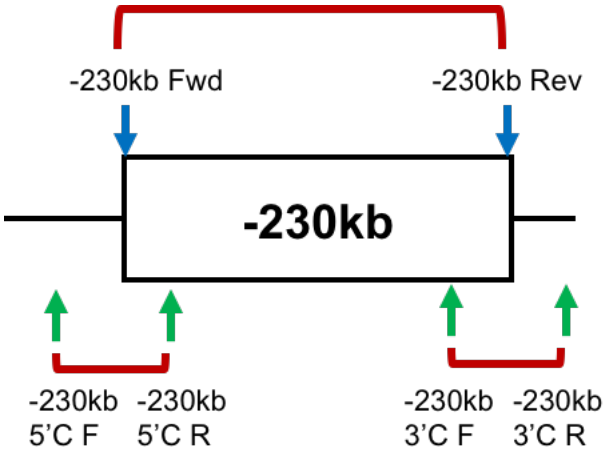


Figure 3.85: schematic representation of strategy used to genotype for excision of the -230kb enhancer in Ccn2Δ230 transgenic line. Blue arrows depict the positioning of the primers used for cloning of the enhancer and green arrows depict the primers used to amplify substrate for *in vitro* CRISPR cutting assay. The red bars depict amplicon that were used for genotyping.

Founder mice were initially genotyped using primers for the -230kb 5' CRISPR cutting assay substrate (Figure 3.86).

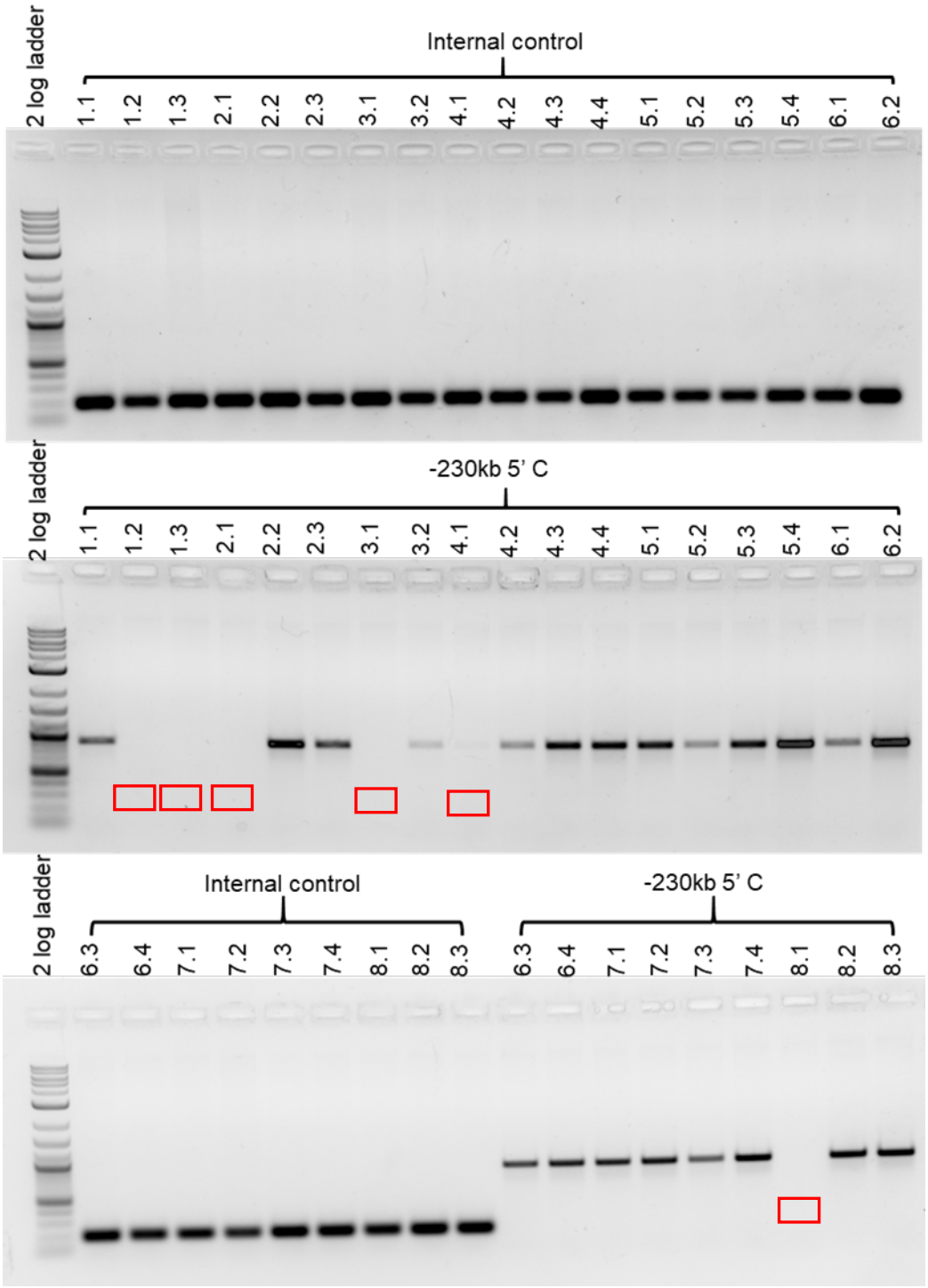


Figure 3.86: Initial genotyping of the Ccn2Δ230 line. Internal control reaction was carried out for each sample, each of which had a strong band. Absence of CRISPR-specific PCR product in mice 1.2-2.1, 3.1, 4.1 and 8.1 (red boxes) suggested possible homozygous knockouts. The faint bands in lanes for 3.2, 4.2, 5.2 and 6.1 suggested potential heterozygosity for enhancer knockout.

This process was repeated using the -230kb 3' CRISPR cutting assay substrate as amplicon of interest (Figure 3.87).

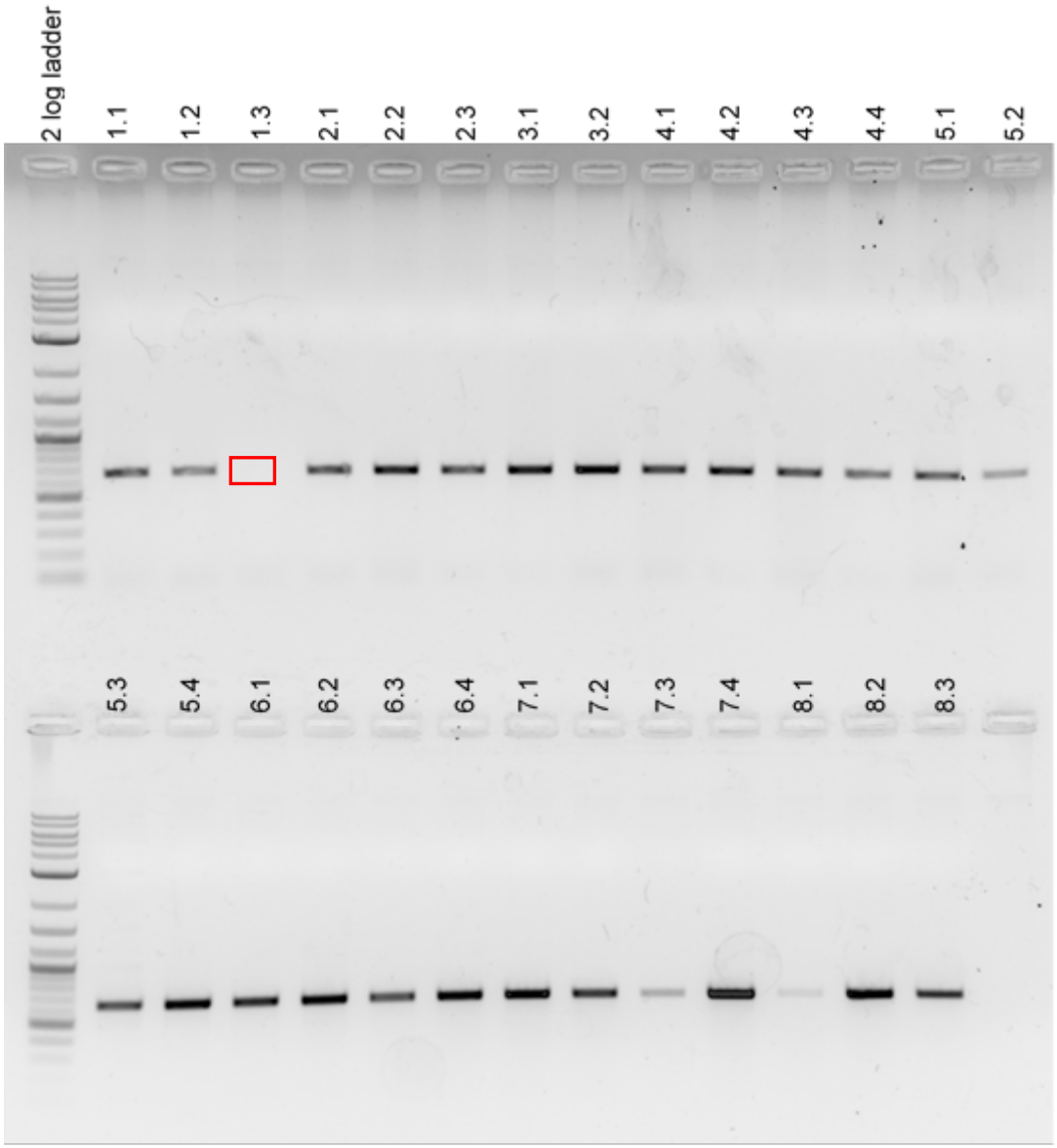


Figure 3.87: Further genotyping of founders from the Ccn2Δ230 line. The -230kb 3' CRISPR cutting assay substrate was used as amplicon (see Figure 3.72). PCR product was formed from each sample excluding 1.3 (red box). Weak bands were observed for 7.3 and 8.1 suggesting heterozygosity for knockout of the enhancer.

From four rounds of genotyping with different amplicon 1.3 and 8.1 were chosen as potential heterozygotes and bred to produce F_1. Resultant offspring were genotyped for the presence of the whole -230kb enhancer, with the expectation that homozygote knockouts would be found with no PCR product on gel (Figure 3.88).

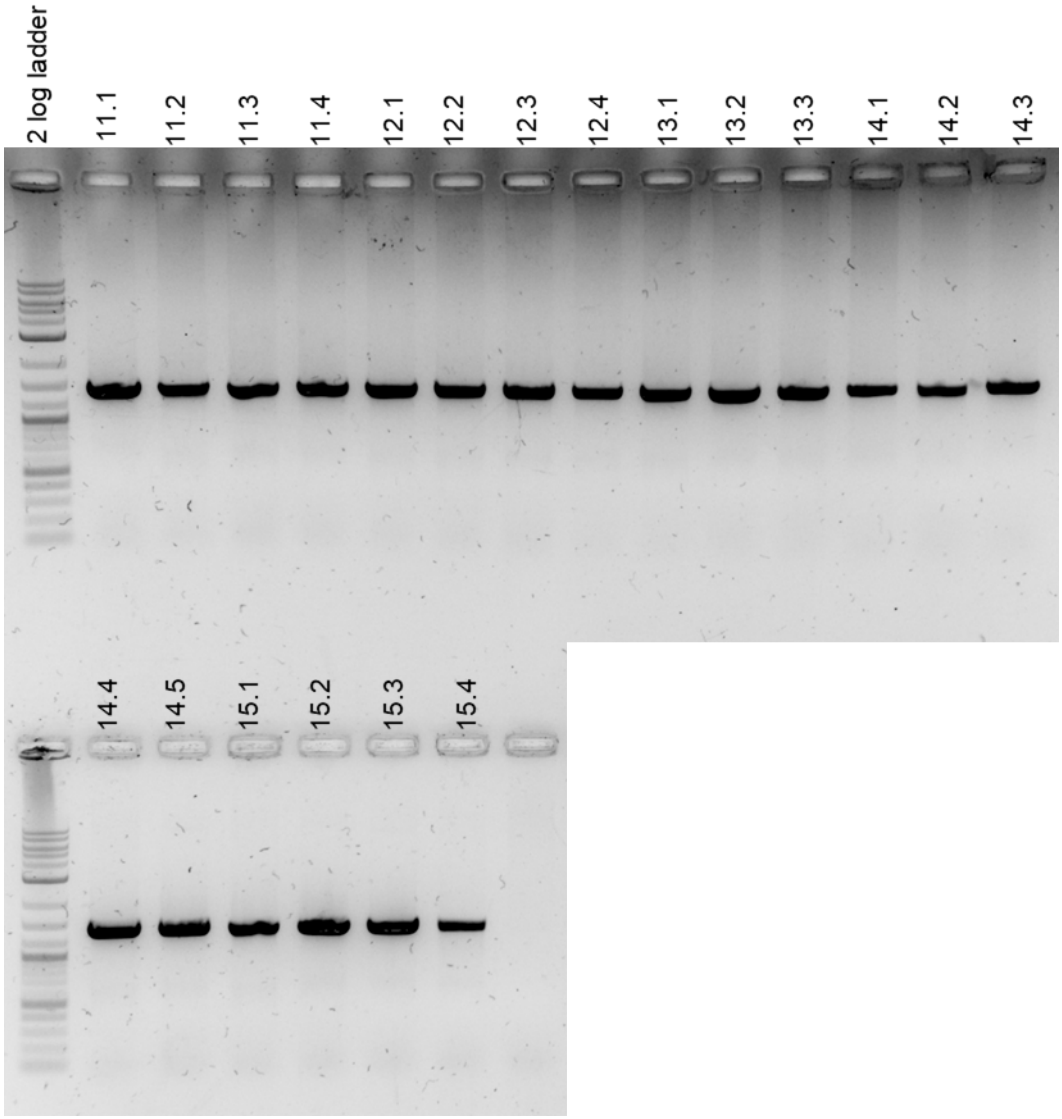


Figure 3.88: Genotyping of F1 Ccn2Δ230 mice. The whole -230kb enhancer was used as amplicon of interest (1.5kb). PCR product was generated in each reaction, illustrating that the enhancer was present in every sample and no mouse was homozygous for deletion of the region.

Given the strong bands obtained for PCR of the -230kb enhancer in F_1 mice, work on this line was not continued as deletion of the enhancer had not occurred.

4. Discussion

4.1 Validation of putative enhancer function

In this thesis, the role of *cis*-acting regulatory elements in regulation of the *Ccn2* gene was investigated. The data presented here demonstrate that novel enhancer regions located -102kb, -137kb, -148kb, -198kb and -230kb upstream of *Ccn2* are each capable of driving reporter gene expression at E15.5 in transgenic mice (Frost *et al.* 2018) (summarised in Chapter 3.6). The -148kb enhancer drove potent chondrocyte based gene expression in both development and adulthood, with a truncated version of this region also displaying enhancer activity at E15.5 and in mature adulthood. Chondrocyte activity was also exhibited by the -137kb and -198kb enhancers, in an articular chondrocyte specific, and more broad manner respectively. This contrasted the function of -230kb and -102kb enhancers which were capable of driving reporter gene expression in osseous tissue and the vasculature respectively. The range of cells and tissues in which β-galactosidase transgene activity was observed are a reflection of cell-type specific TF engagement with the enhancer sequences. These interactions therefore facilitated enhancer function which culminated in chromatin organisation that mediated transcriptional output from the *LacZ* reporter gene. Moreover, each of the tissues in which β-galactosidase activity was visualised have previously been reported to express *Ccn2* endogenously (Friedrichsen *et al.* 2003, 2005; Ivkovic *et al.* 2003). Therefore, *enhancer activity reflected regulation of Ccn2 transcription in a highly temporospatial specific manner*. This echoes many other publications that have associated *cis*-regulatory elements with cognate genes using this methodology and ethos (Li *et al.* 2018; Nord 2015; Pennacchio *et al.* 2006). Furthermore, limb specific enhancers have been well established as fundamental regulators of skeletogenesis (Cotney *et al.* 2012; Infante *et al.* 2015; Kvon *et al.* 2016).

4.1.1 Further *in silico* based evidence supporting enhancer function

Further evidence to ratify the enhancer regions predicted in the current work has come from recently released ENCODE datasets in the mm10 Genome Browser Build. Upon LiftOver of the 300kb region of interest and putative enhancers identified in the mm9 Genome Browser build in the current study, the accuracy of enhancer predictions made in the current study can be further scrutinised. Firstly, the ‘ENCODE Enhancer-like regions mm10’ track set available as part of the

ENCODE Encyclopaedia version 3 at: http://zlab-annotations.umassmed.edu/enhancers/. This allows the prediction of 'enhancer-like' genomic regions at E11.5 and E14.5 in limb based on H3K27ac and DNase I sensitivity (yellow bars, Figure 4.1). Enhancer-like regions predicted at E11.5 limb in this data track concur with the -255kb, -137kb and -4kb enhancers identified in the current work. This ENCODE track also has prediction of enhancer-like regions at E14.5 overlapping with the -255kb, -230kb, -102kb and -4kb enhancers identified in the current work. Interestingly, the -198kb enhancer which drove transgene activity in three founder embryos at E15.5 in the current project, was not predicted to have enhancer-like activity at E14.5, which is unexpected given that based on the datasets used in the mm9 browser, the peaks for DNase I sensitivity and H3K27ac were comparable to that of -102kb which was given enhancer-like predicted function in the new ENCODE data set (Figure 3.1). At E14.5, peaks for H3K27ac and therefore active regulatory elements concur with the position of each putative enhancer described in the current project. Peaks for this modification were greater at E14.5 compared to E11.5, suggesting that at E11.5 enhancers may be poised, with greater chromatin accessibility and therefore enhancer activity as embryonic development proceeds (Rada-Iglesias *et al.* 2011).

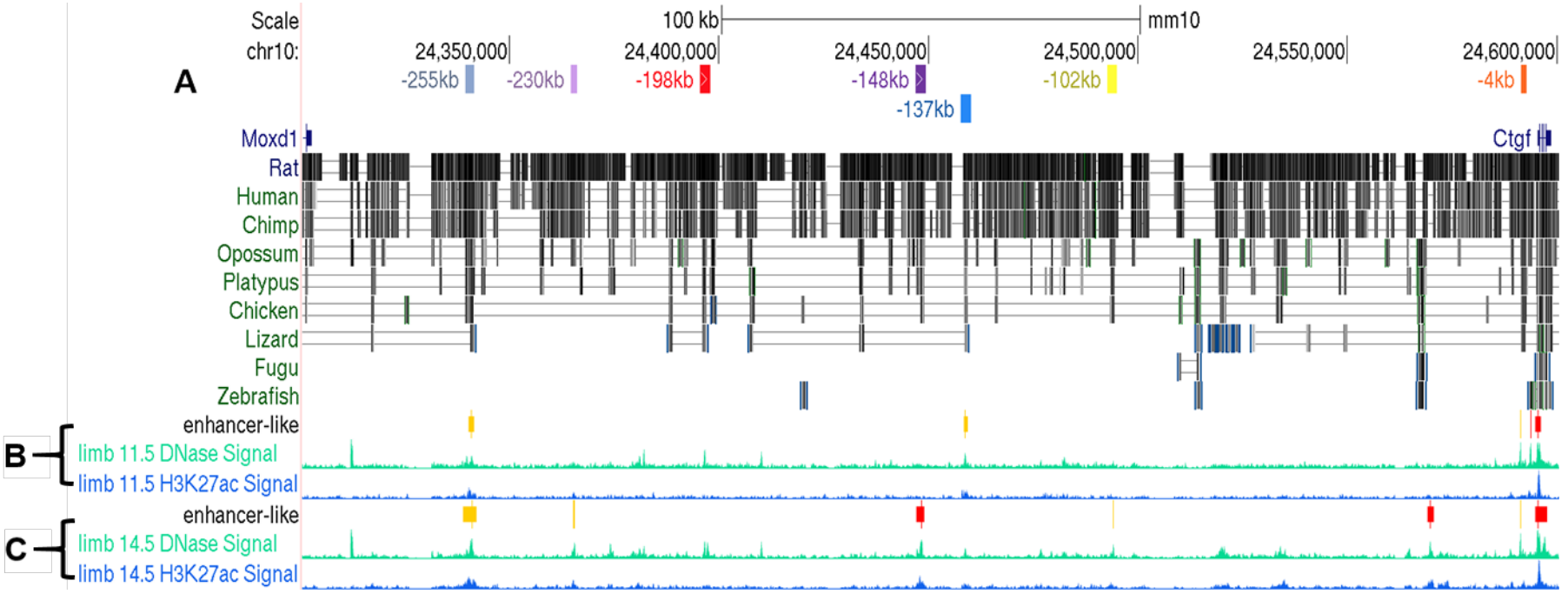


Figure 4.1: UCSC Genome Browser graphic of ENCODE enhancer-like regions upstream of *Ccn2* compared to positioning of enhancers described in the current study. This dataset plots DNase I hypersensitivity and H3K27ac at E11.5 (B) and E14.5 (C) in order to predict enhancer regions Yellow bars represent predicted enhancers, whereas red bars represent predicted promoter regions.. Overlay of this track with the putative enhancers of *Ccn2* (A) allows further validation of enhancer predictions.

Enhancer predictions are also annotated in the recently available candidate Regulatory Element (cRE) resource from the ENCODE consortium. This track set is available through the Search Candidate *cis*-Regulatory Elements (SCREEN) interface within the ENCODE Encyclopaedia version 4 (available at: http://screen.encodeproject.org/) (The ENCODE Project Consortium 2011, 2012). This track set is based on data for H3K4me3 and H3K27ac presence in addition to DNase I hypersensitivity. Within limb tissue samples, each enhancer predicted in the current study is predicted to be a cRE. As with the enhancer-like track, regulatory function is acquired after E11.5 for the -230kb and -148kb regions. However, this contrasts the findings of enhancer function from the CTGF148 transgenic mouse line utilised in the current study, in which chondrocyte based enhancer activity was observed at E12.5. This enhancer was also active in transgenic founders at E11.5.

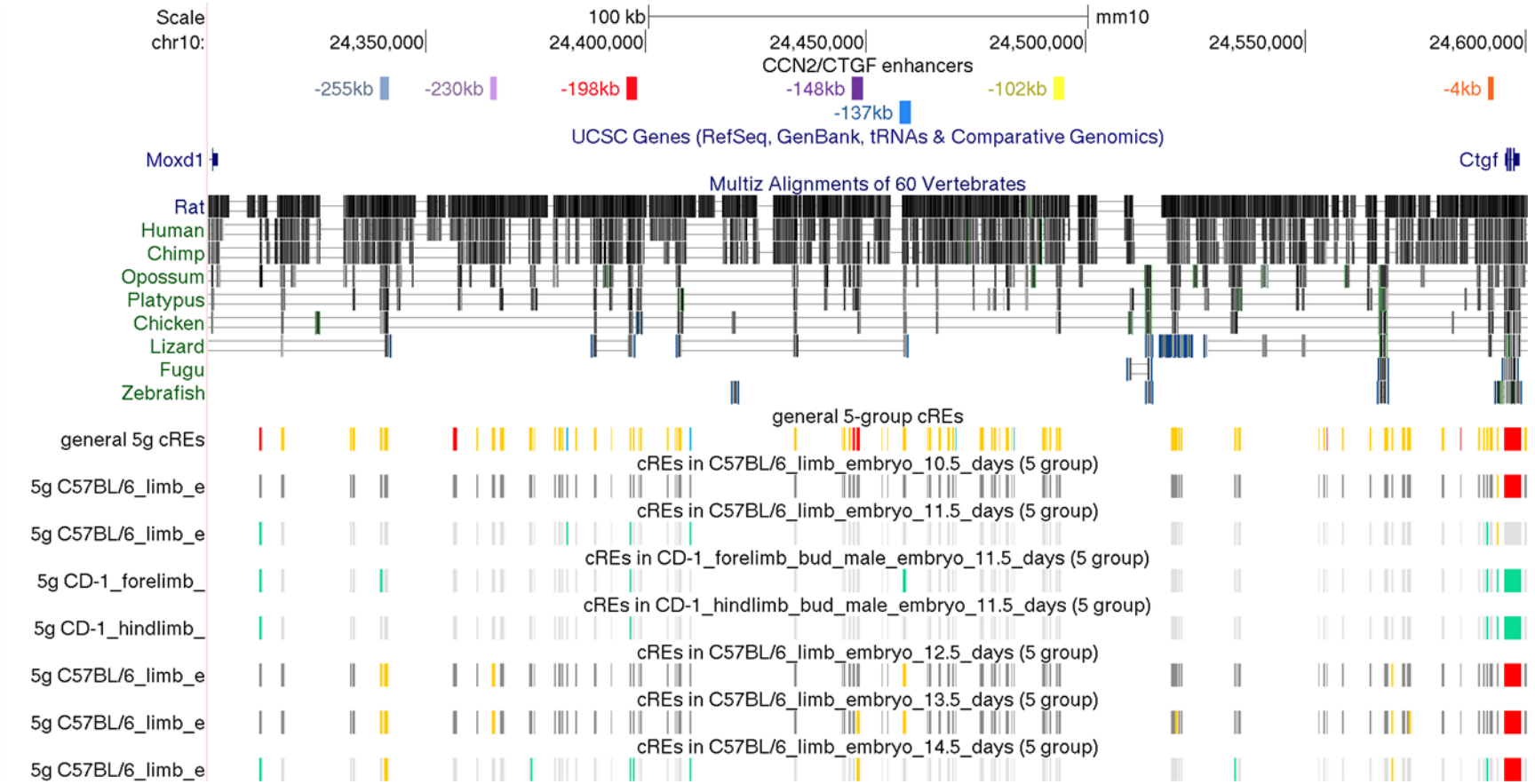


Figure 4.2: cRE data set compilation in mm10 UCSC Genome Browser compared to enhancer positions examined in the current study. These samples encompass different time points from E10.5 to E14.5. Red blocks denote promoter predicted activity, yellow blocks for enhancer-like activity and green blocks for open chromatin hypersensitive to DNase I. The promoter region for *Ccn2* (*Ctgf*) is prominently marked with red bars at each time-point. There is also a red promoter-associated block within the -148kb enhancer region

A major oversight in these two ENCODE enhancer prediction track sets is the lack of data concerning H3K4me1 and H3K27me3 modification in samples. H3K27me3 would allow the distinction of repressive chromatin state, and therefore poised enhancers, or those that may be active in a different cellular context (Zhu *et al.* 2013b). The use of H3K4me1 data would be a much more powerful tool in the prediction of the enhancer regions, as it would allow comparison of promoter and enhancer function, as carried out using datasets in the mm9 Genome Browser used in the current study. The lack of H3K4me1 data is puzzling given the strong consensus that this modification is important in enhancer prediction (Creyghton *et al.* 2010; Rada-Iglesias *et al.* 2011). H3K4me1 is much stronger than H3K4me3 at enhancers, with the latter modification being very weakly observed, if at all within enhancers (Heintzman *et al.* 2007). In addition, these ENCODE tracks suggest that -148kb is a promoter (red blocks Figures 4.1 and 4.2) based on H3K4me3. In the mm9 Genome Browser track set used in the current study, there is a negligible peak for H3K4me3 and substantial peak for H3K4me1, thereby suggesting enhancer function rather than promoter activity. Moreover, enhancers may be distinguished from promoter regions of the basis of the ratio of H3K4me1 to H3K4me3 (Calo and Wysocka 2013). The H3K4me3 peak at the true *Ccn2* promoter is much richer than in -148kb in each of the *in silico* resources, these factors therefore suggest that the -148kb region is an enhancer rather than a promoter. However, promoter function cannot be completely ruled out, in line with the findings of Mikhaylichenko *et al*. (2018) who found that enhancers may function as weak promoters.

However, a further convoluting factor in the function of -148kb is that transcription is initiated from within the -148kb enhancer. The 'Ensembl gene predictions' track within the UCSC Genome Browser shows that a long intergenic non-coding RNA (lincRNA), ENSMUST00000134627 is transcribed from within -148kb (Hubbard *et al.* 2002) (highlighted blue box, Figure 4.3). As aforementioned in Chapter 1.3.3.5, enhancers may be transcribed to produce eRNA. ENSMUST00000134627 has not been characterised, but it is possible that it is the eRNA product of -148kb rather than a lincRNA (Kim *et al.* 2015). None of the other enhancers identified within the current study are co-localised with the location of lincRNA (Figure 4.3). Beyond the scope of the current project, future work examining the expression of this transcript would be pivotal in understanding -148kb. Most important would be ascertainment as to whether the transcript produced near this locus is eRNA, and if so, how does it function in the activity of the -148kb enhancer. This could occur through eRNA

interaction in chromatin looping or transcriptional elongation at *Ccn2*. Furthermore, assay of ENSMUST00000134627 and *Ccn2* expression in various tissues and at different time points could enable refinement of the understanding of the specific temporospatial circumstances in which the enhancer functions. *In vivo* findings suggest that the enhancer functions most potently in chondrocytes (Chapter 3.4). This idea would be reaffirmed if the transcript was most abundant in chondrocyte cells and cartilaginous tissue. This therefore leads to the question as to whether ENSMUST00000134627 participates in enhancer activity, or its location is a coincidence with the lincRNA functioning in a completely different process unrelated to the enhancer? However, the fact that the -148kb full and short transgenic construct versions of the enhancer were able to drive reporter gene expression would suggest that ENSMUST00000134627 is not essential in its function. ENSMUST00000134627 is produced from approximately 600bp into the enhancer sequence through to approximately 3.2kb downstream of the 3' enhancer boundary. Reporter gene constructs contained the enhancer alone spanning from 5' boundary to 3' boundary of the enhancer sequence and therefore the full ENSMUST00000134627 transcript could not be transcribed. This would suggest that the RNA is not necessary in the function of the enhancer, as the enhancer sequence alone drove reporter gene expression. Moreover, the -148kb short transgenic line construct did not contain the start site for the ENSMUST00000134627 transcript. These findings highlight the need for further examination of the enhancer and the mechanism through which it increases transcriptional output.

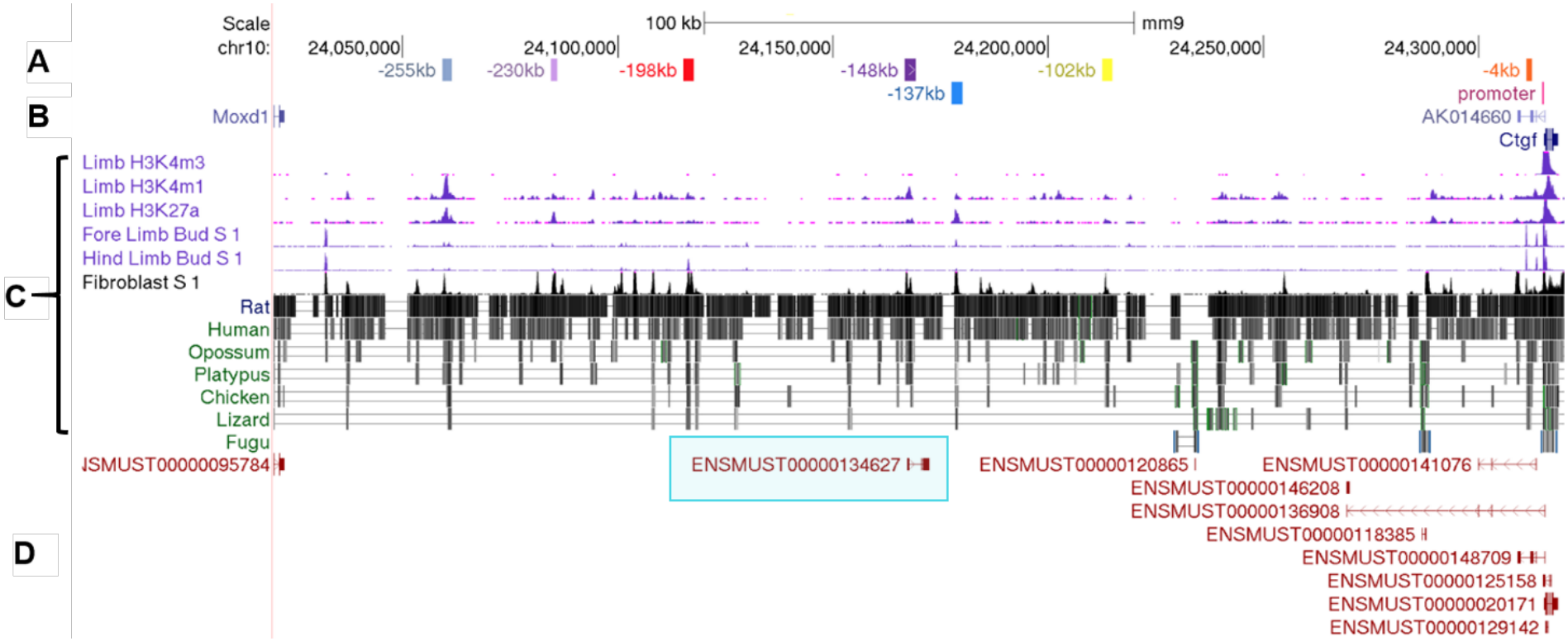


Figure 4.3: UCSC Genome Browser graphic of transcripts occurring within region upstream of *Ccn2*. Transcripts from the ENSEMBL Gene Prediction Track (D) are produced bi-directionally from the promoter for the gene. The transcripts produced in a downstream to upstream direction are probably uaRNA that are a consequence of RNA pol II occupation at the promoter, with no protein coding potential. The only transcript produced near any enhancer is ENMUST00000134627 which is transcribed from within the -148kb enhancer (highlighted with blue box).

Predictions of enhancer function can also be bolstered by transcription factor-DNA interaction datasets. Moreover, Dogan *et al*. (2015) recommend that TF interaction with DNA sequence is a more reliable indicator of enhancer function than histone modification alone, with the suggestion that false predictions may be made using chromatin modification datasets only. The findings of Kwasnieski *et al*. (2014) support this with activity demonstrated in 26% of enhancers predicted using ENCODE dataset histone modifications, with the suggestion that TFBS are a more accurate tool in predicting enhancer activity. Of course, a major consideration in the use of ChIP-Seq datasets is the context in which the binding of protein to DNA was assessed i.e. the cell type or tissue assayed. This must be considered in drawing conclusions about the temporospatial context in which an enhancer functions and the cell signalling mechanisms that underpin this. For example, false negative relationship between an enhancer and TF may be observed regarding an enhancer that is only active during embryonic development yet a ChIP-Seq sample from adult tissue is used to look for TF-enhancer binding. Furthermore, a major attribute of enhancers is the plasticity in TF binding dependent on cellular context, with the change in chromatin state and transcription factor occupancy over time dictating refined gene transcription in a highly specific temporospatial manner (Creyghton *et al.* 2010; Spurrell *et al.* 2016).

The assay of histone acetyltransferase p300 binding to DNA has been shown to be an effective tool in the identification of enhancer regions, with Visel *et al*. (2009) using ChIP-Seq of p300 to predict enhancer regions that were subsequently validated *in vivo*. Although not carried out during the prediction of enhancers for the current project, p300 can nonetheless provide further evidence for the enhancer function of the *cis*-acting regulatory regions described herein. Application of p300 ChIP-Seq in mouse embryonic fibroblast cells (MEFs) dataset (GEO DataSet GSM2417166, Chronis *et al*. (2017)) to the intergenic region that was the focus of the current project serves as further validation of the enhancer regions described in, as demonstrated in Figure 4.4. There is a peak for p300 interaction within each of the putative *Ccn2* enhancers. In addition, although peaks are spread throughout the 300kb region upstream of *Ccn2*, they are more concentrated in the vicinity of the enhancers identified herein.

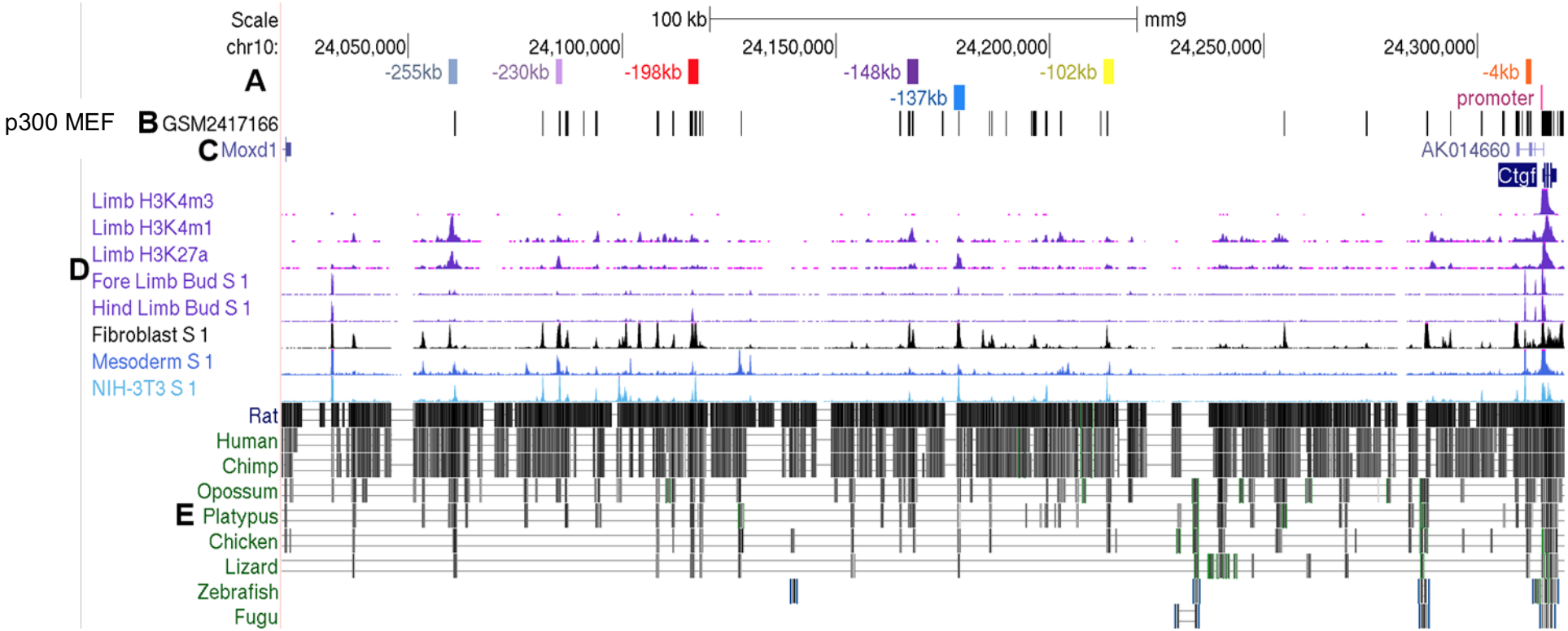


Figure 4.4: ChIP-Seq validated binding of p300 within the intergenic region upstream of *Ccn2*. GEO DataSet GSM2417166 (Chronis *et al.* 2017) regarding ChIP-Seq validated interaction between p300 histone acetyltransferase and DNA in mouse embryonic fibroblasts cells, incorporated into the 300kb region upstream of *Ccn2* (B and C) provides many peaks that illustrate interaction between p300 and the genomic sequence. Many of the peaks fall within enhancers of *Ccn2* described in the current project (A). Peaks are aligned with enhancer associated chromatin characteristics (D) and evolutionary conservation of sequence (E)

The accuracy of *in silico* predictions is drawn into question with the -255kb putative enhancer. In the mm9 Genome Browser session used to predict enhancers (Figure 3.1 and D), the region was highly conserved and the peaks for H3K4me1 and H3K27ac were substantially larger for -255kb compared to those of the other putative enhancers. There is also a peak for p300 within the region (Figure 4.4). In addition, this region is the only annotated to function at both E11.5 and E14.5 in the enhancer-like ENCODE track and is predicted to function from E12.5 in the cRE track (Figures 4.1 and 4.2 respectively). It would therefore be expected that this region would function as a strong enhancer *in vivo*, driving high levels of reporter gene expression in the transgenic model. This however was not the case, with no positive staining for β-galactosidase activity in any founder. This was also true for the -4kb putative enhancer. Although beyond the scope of the current project due to time and fiscal constraints, more experimental repeats of the creation of transgenic mice harbouring the enhancer driving reporter gene expression should be conducted before the -255kb and -4kb regions are ruled out as enhancers.

4.1.2 -102kb functions as an enhancer in the vasculature

The only previous attempt to examine enhancers within the region upstream of *Ccn2* was made as part of the VISTA Enhancer Browser (Visel *et al.* 2007b). In this study, the only enhancer of *Ccn2* that was posited concurred with a 1680bp fragment within the 2282bp -102kb enhancer described herein. The VISTA group examined this region at E11.5 in founder reporter transgenic mice and found it to be non-functional. In the current study, the -102kb enhancer functioned in a highly reproducible manner within the superficial vasculature at E15.5 (Chapter 3.2.2). This suggests that the enhancer is inactive at E11.5, and activated by E15.5, or that the full sequence of the enhancer is required for function. This latter idea is not likely as the most conserved part of the sequence, and region with histone modification peaks sits in the middle of the enhancer and were present in both transgenic constructs.

The localisation of -102kb function tallies with previous studies that have demonstrated *Ccn2* expression within the vasculature (Friedrichsen *et al.* 2003). Shimo *et al*. (1999) demonstrated that CCN2 functions in the proliferation, migration and adhesion of vascular endothelial cells. Hall-Glenn *et al*. (2012) used transgenic mice harbouring *LacZ* under control of the -4kb to TSS of *Ccn2* (outlined in Chapter 1.4). This group observed β-galactosidase activity throughout the vasculature at

E13.5. Moreover, this included the dermal microvasculature; akin to the pattern observed within the current study. A broader 160kb genomic fragment controlling GFP expression within the BAC system (described in Chapter 1.4) drove transgene expression within both vascular smooth muscle cells within large arterial vessels in addition to endothelial and mural cells of the microvasculature (Hall-Glenn *et al.* 2012). An important finding from this work was that microvasculature of *Ccn2* embryos was aberrant with defective vessel maturation. Endogenous function of the -102kb enhancer therefore may act to reinforce microvasculature based expression of *Ccn2* and subsequent vessel maturation. The findings of Hall-Glenn *et al*. (2012) were observed within the microvasculature of several tissues later in embryonic development than tested in the current study. It would therefore be useful to further examine -102kb activity at more time-points than E15.5, with the possibility that the enhancer continues to function as microvasculature vessels mature.

Upon inputting the -102kb sequence into the TRAP database, several transcription factors with vascular or endothelial function are predicted to bind within the enhancer (Chapter 5.5). There are several consensus binding sites for MEF2, c-MYC, ETS2, GATA and HIF2α within the -102kb enhancer. Each of these transcription factors is important in endothelial cell function and angiogenesis (De Val and Black 2009; Skuli *et al.* 2009; Wei *et al.* 2009). Beyond the scope of the current project, further work is required in order to validate the interaction between these TF and cognate sequence within -102kb. Nonetheless, the TRAP results do support the finding from the *in vivo* reporter gene assay that -102kb primarily functions in endothelial cells during angiogenesis and the development of the vasculature. Future work should validate this with EMSA experiments examining the capacity of transcription factors to bind to these sequences.

4.1.3 -137kb enhancer functions in articular cartilage

Highly reproducible transgene expression was also driven by the -137kb enhancer *in vivo* (Chapter 3.2.3). Chondrocytes were the only cell type in which transgene activity was observed at E15.5, with the enhancer only functioning within a specific sub-population of chondrocytes in proximity to articular joints. This occurred most notably within the elbow, knee, wrist and ankle. In addition, X-gal staining was reduced with distance from the joint surfaces (Chapter 3.2.3, Figures 3.6 and 3.7). There was no β-galactosidase activity within the stylopod articular chondrocytes of

the shoulder and hip. Interestingly, *in silico* prediction of enhancer function suggests that the -137kb enhancer is active at E11.5, with reduction in activity and open structure of chromatin by E14.5 (Figure 4.1). The temporal organisation of limb development with regard to proximal to distal advancement of differentiation, combined with these findings suggest that the enhancer functions in a highly temporospatial manner to drive the expression of *Ccn2* during joint formation. In addition, the change in X-gal staining intensity with distance from the joint tallies with models of joint development where early joint cells are pushed into the endochondral epiphysis over time (Shwartz *et al.* 2016). Moreover, a limitation in the use of *LacZ* reporter genes is the half-life of the β-galactosidase protein, with the enzyme persisting in a cell after the enhancer regulating its expression has been, and is no longer active (Bi *et al.* 1999). Therefore, -137kb may have functioned more potently at an earlier time-point in the early interzone and articular chondrocytes of the shoulder and hip joints, but using the E15.5 time-point of the current study this activity would not be observed. Moreover, in ChIP-Seq datasets pertaining to active regulatory histone modification H3K27ac within the limb at E11.5 (Infante *et al.* 2015) (GSE64055), -137kb is the only enhancer that was validated in the current project to contain peaks for interaction and therefore the regulatory function associated histone modification at this time point (Figure 4.5).

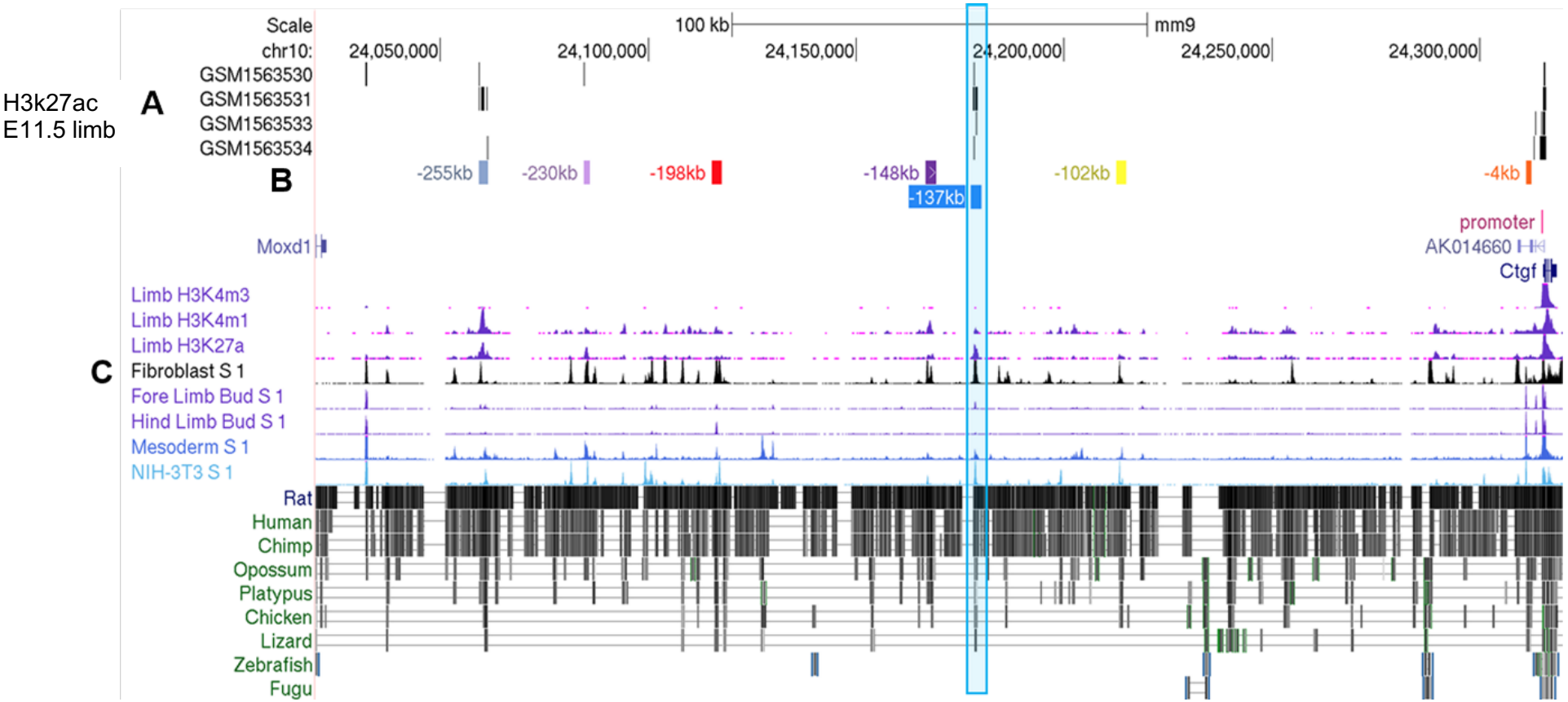


Figure 4.5: UCSC Genome Browser visualisation of ChIP-Seq datasets for H3K27ac in E11.5 limb. Import of H3K27ac ChIP-Seq datasets of limb tissue at E11.5 (Infante *et al.* 2015) (A) enable refinement of the estimation of enhancer function earlier than the E15.5 time-point used in reporter gene assay in the current study (B) . Across three tissue samples there are peaks for H3K27ac and therefore active regulatory function in the -137kb enhancer and *CCN2* promoter. The -148kb enhancer functioned in vivo at E11.5 but not strongly within limb tissue which may explain why there is no peak for H3K27ac in any sample (A). The greater number of peaks for H3K27ac at E14.5 (C) suggests the chromatin is remodelled to acquire regulatory function between E11.5 and E14.5

Enhancer function specific to articular tissue has been demonstrated in previous studies, such as that of Chen *et al*. (2016) who describe *cis*-regulatory elements that drive expression of *GDF5* within specific sub-populations of chondrocyte and joint cells. This study found discrepancy in enhancer function between proximal and distal joints, as observed in the current work. GDF5 could be a regulator of *Ccn2* expression within the articular joint through the function of the -137kb enhancer. Furthermore, the pattern of the β-galactosidase expression pattern driven by the enhancer is similar to GDF5-related *LacZ* expression that has been observed in several previous studies (Chen *et al.* 2016; Koyama *et al.* 2008; Shwartz *et al.* 2016). Transcription factors that function downstream of GDF5 such as Smad1/5 are not predicted to bind within the -137kb enhancer upon input of either the full enhancer or short region sequence into the TRAP tool (Chapter 5.5). The lack of enhancer function within the joint cavity tallies with the expression of SOX9 which is not found within developing joints after E14.5 (Soeda *et al.* 2010). Moreover, the SOX9 consensus binding site validated by EMSA in the current study could be utilised by SOX9 in regulating *Ccn2* transcription during early articular joint formation. Therefore, as SOX9 expression is lost within this region as development proceeds, enhancer function would also decrease. Of course, SOX9 expression is retained throughout the endochondral anlage until the hypertrophic zone, and enhancer function was only found near to the articular zone, suggesting that further transcription factors are required in order to direct enhancer activity that is specific to articular chondrocytes within this narrow differentiation period. This highlights the complexity in the regulatory interactions that underpin joint formation and the need for further elucidation of -137kb activity within this process. Future work should seek to define how activity of the enhancer changes throughout joint development. Principle in this would be application of -137kbHsp68LacZ construct used herein to further developmental time-points such as E11.5 and E13.5 whereby the role of the enhancer within joint would be better characterised.

The -137kbshortHsp68LacZ construct could also be examined in this process. Our attempt to create transgenic mice harbouring this construct was unsuccessful in this project. Repetition of this process with more founder animals would enable determination of how much of the putative sequence is required for the function of the enhancer. Another important facet of future study would be further examination of the TF that facilitate enhancer function, thereby allowing better understanding of the signalling mechanisms that dictate joint formation and the role of *Ccn2* in this process.

4.1.4 The -148kb enhancer drives gene transcription within chondrocytes

The most prevalent enhancer activity *in vivo* was driven by the -148kb enhancer (Chapter 3.3-3.5). In endochondral ossification, -148kb functioned in hypertrophic chondrocytes in each reporter transgenic founder at E15.5. The creation of the CTGF148 line enabled greater characterisation of ability of the enhancer to regulate gene transcription at several time points. The most potent β-galactosidase activity occurred within cartilaginous tissue during both embryonic development and adulthood (Chapter 3.4). This tallies with the need for CCN2 expression within cartilaginous tissue and the severe phenotype resulting from ablation of *Ccn2* expression in chondrocytes (Ivkovic *et al.* 2003). -148kb functioned strongly in hyaline cartilage on a global scale in elements such as the costal cartilage and articular cartilage from cartilage anlage in development through to established cartilage in mature adulthood. Transgene activity was also recorded in fibrocartilage, such as within the intervertebral disc. Whilst the whole endochondral growth plate is transgene positive at E15.5, the chondrocyte specific nature of enhancer function from this time-point onwards is indicated by the stratification of activity as ossification is advanced at E17.5, with loss of transgene activity in osseous cells and tissue. Indeed, as aforementioned, the transgene activity may reflect enhancer function at an earlier time-point due to β-galactosidase protein half-life.

In embryonic development, the results of both Friedrichsen *et al*. (2003) and Ivkovic *et al.* (2003) demonstrate endogenous *Ccn2* expression within cartilage globally, such as within the chondrocranium, ribs, vertebrae, limbs, and digits, matching the patterns observed for -148kb activation during embryonic development. Enhancer activity within the Meckel's cartilage recapitulated the work of Shimo *et al*. (2004) who demonstrated endogenous expression of Ccn2 within this tissue during embryonic development. Much of the work concerning CCN2 within cartilage has not focused on characterising the expression profile of the gene and protein at several time-points, but more commonly *ex vivo* study of cartilage explants. However, CCN2 expression has been observed in adult articular cartilage, akin to the transgene in the current study (Kubota and Takigawa 2011).

Transgene expression within adult articular cartilage was also demonstrated by a truncated version of the -148kb enhancer in the -148kb_short D transgenic line

variant. In this line variant and for founder A progeny, the enhancer also drove chondrocyte and perichondrium based transgene expression at E15.5, akin to the endogenous expression of *Ccn2* (Ivkovic *et al.* 2003). These findings suggest that chondrocyte based transcription is directed from an element within this truncated region, but the full enhancer sequence is required for increased transcriptional output. Moreover, the variability of transgene expression in several tissues suggests that specificity in enhancer function directing expression in a temporospatial manner was reduced. A pertinent question is whether the enhancer sequence of approximately 1kb removed in the creation of -148kb_short would function *in vivo* if a *lacZ* reporter construct was created for this variant and tested in transgenic mice. This would allow further refinement of the understanding of the function of the sequences within the enhancer.

4.1.5-198kb enhancer function

The variety in reporter gene expression pattern and low number of founder embryos generated in the study of -198kb function hampered attempts to characterise the role of the enhancer, and tissue-type in which it functions most potently. If the expression pattern of founder A is true, with strong enhancer function within hypertrophic chondrocytes, it would suggest that the -198kb region would function as a further enhancer of *Ccn2* transcription within chondrocytes. Whilst beyond the means of the current study, further repeats of -198kbHsp68LacZ assay at E15.5 are required in order to validate this.

4.1.6 The -230kb enhancer drives gene transcription within osseous tissue

Transgene expression driven by the -230kb enhancer was observed within osseous tissue. There was however some variation in the time point that the enhancer was active within the two founder variant progeny. At E15.5 founder A gave rise to potent *LacZ* expression within periosteal tissue and within the primary ossification centres of long bones. The progeny of founder C did not exhibit transgene expression at E15.5, yet primary spongiosa osteoblasts were positive for β-galactosidase activity at P7 and six weeks before repression in mature adulthood. The osteoblastic nature of expression suggests that TF involved in bone cell lineage commitment are predominant in the regulation of the function of -230kb, bestowing osteoblast specificity in enhancer activity and therefore *Ccn2* transcription. This would follow further studies that have demonstrated osteoblastic

expression of *Ccn2* (Arnott *et al.* 2011). However, further validation of this is required through the *in vivo* assay of enhancer function in founder transgenic reporter mice such as with repetition of -230Hsp68LacZ line generation. Further important information would have also been gleaned through the assay of β-galactosidase activity in Founder A in adult, however tissue samples were not available in order to allow this.

4.1.7 *In vitro* validation of enhancer function

Discrepancy between the modelling of enhancer activity and *in vivo* validation of function is highlighted by comparison of the results from *in vitro* and *in vivo* assay of enhancer capacity to drive reporter gene expression. This was exemplified by the fact that the -4kb and -255kb regions drove luciferase reporter gene expression *in vitro* yet neither demonstrated function in any *in vivo* assay that was carried out. This could however be due to sample size or function at a different time point.

Moreover, the reporter gene expression induced by both of these regions *in vitro* was greater compared to that of the enhancers proven to function *in vivo*. Ultimately, the results of *in vitro* reporter gene assays described herein demonstrate that the enhancer regions can drive gene expression in tandem with the *Ccn2* promoter region, but do not aid in understanding the regulatory interactions that facilitate enhancer function. Episomal DNA in reporter plasmids can acquire chromatin characteristics, and therefore be a better representative of the endogenous enhancer than plasmid DNA alone (Catarino and Stark 2018). However, the TF that dictate enhancer function may not be present in immortalised cell lines. For example, the HTB-94 chondrosarcoma line used herein has been shown to exhibit limited similarity in gene transcriptional profile with primary articular chondrocytes (Gebauer *et al.* 2005). Expression of *Sox9* within TC28-i2 cells have been used to demonstrate the effectiveness of this cell type in modelling chondrocyte function, yet primary chondrocytes are still posited as a better model (Finger *et al.* 2003). In the current study, use of the SOX9 TFBS containing -137kb and -230kb enhancer regions within TC28-i2 cells did not lead to marked changes in transcriptional output compared to the use of the *Ccn2* promoter alone.

Future attempts to study function of *Ccn2* enhancers within a chondrocyte context *in vitro* should utilise primary chondrocytes or the ATDC5 cell line. Primary chondrocytes most accurately reflect *in vivo* chondrocyte behaviour. ATDC5 cells

can be stimulated to undergo chondrogenic differentiation and are well regarded as an *in vitro* model of endochondral ossification (Yao and Wang 2013). During differentiation, ATDC5 cells express TF and ECM proteins in a similar manner to chondrocytes *in vivo* (Akiyama *et al.* 2000; Chen *et al.* 2005). Enhancer activity could be assayed at several time points in ATDC5 differentiation in order examine if enhancer activity changes as the prevalence of cell lineage-specific transcription factors such as SOX9 and RUNX2 change. For example, it would be expected that activity of the -137kb and -148kb enhancers would decrease with change to osteoblastic cell prevalence, whereas the activity of -230kb would increase in this late stage of differentiation. Beyond the time-frame of the current project, further work should also focus on the effect of manipulating cell signalling pathways on enhancer activity. For example treatment with cytokines such as IL-1β, TGF-β or knockdown of protein expression using siRNA would enable further understanding of the key TF-DNA interactions and the context in which they occur in controlling enhancer activity.

4.1.8 *In vivo* function most accurately reflects endogenous enhancer activity

All of these findings highlight the fact that there are some limitations with the annotation of enhancer function using *in silico* and *in vitro* approaches; however, they are nevertheless useful in the preliminary stages of enhancer identification (Catarino and Stark 2018). Many predicted enhancers do not function *in vivo* (Pennacchio *et al.* 2006), thus *in vivo* assay of reporter expression driven by enhancers gives the most faithful recapitulation of endogenous enhancer activity. *In vivo*, the array of signalling mechanisms and transcription factors are endogenous and correspond to a specific cell type at a defined time-point. Moreover, the enhancer function within a cell type, tissue and whole organism can be assessed in one assay (Kvon 2015; Nord 2015). The complex signalling mechanisms that underpin the behaviour of a cell within its physiological tissue environment is lost *in vitro*, thereby limiting the ability to understand enhancer activity within immortalised cell lines.

However, the *in vivo* system employed herein is still error prone in the assessment of enhancer function due to transgenic construct site of integration within the genome. The constructs were not directed to specific genomic loci during integration, and therefore landed within an uncontrolled and unknown portion of the genome (Nord 2015). Transgenic enhancer activity therefore does not reflect the

chromatin landscape in which the enhancer is found endogenously and the interactions that underlie chromatin topology in enhancer function (Catarino and Stark 2018). Moreover, regulatory elements from other genes could have a bearing on enhancer activity resulting in non-specific gene expression. This therefore highlights the need for many replicates of founder transgenic animals in order to ascertain reporter gene expression that is true to the function of the enhancer (Pennacchio *et al.* 2006). This is also emphasised by the fact that in the current study, some embryos and adult mice were genotyped as being positive for transgenic construct, yet no transgene function was observed, contrasting other founders bearing the same construct in which an enhancer was active.

4.2 What regulates enhancer function in skeletogenesis?

Given the complicated hierarchy of transcription factors that bind to any enhancer, it would be expected that several lineage-specific TF bind with co-activators in order to refine enhancer activity in a highly specific temporospatial manner. However, identification of TFBS within the enhancers does allow some refinement of the understanding of the signalling mechanisms that dictate and regulate enhancer function. Ultimately, future mutation of key TFBS within each enhancer and analysis of the impact on enhancer function will allow the identification of the lynchpin sequences that are fundamental in enhancer activity and therefore *Ccn2* transcription.

4.2.1 SOX9 is capable of interacting with -137kb, -148kb and -230kb

As stated earlier, ChIP-Seq is an important method in the study of enhancer regions. Aside from the identification of enhancer regions, this technique can also be used in the characterisation of enhancer regions and the regulatory mechanisms that control function. Ultimately, the work in the current project would have been greatly improved through the application of ChIP-Seq in order to further examine and validate SOX9 TFBS and their utilisation within chondrocytes. Nevertheless, publicly available ChIP-Seq datasets were useful in the identification of TF consensus binding sites that were subsequently examined using EMSA, and can be used to speculate as to which signalling mechanisms regulate enhancer function.

As mentioned in Chapter 1.1.2, He *et al*. (2016) hypothesise that synergistic AP-1 –SOX9 interaction is important in driving chondrocyte hypertrophy. The data generated in this study regarding AP-1 family member Jun binding within the genome of P1 derived rib chondrocytes is publicly available as part of the GEO DataSets repository: GSM1891979 (He *et al.* 2016). This data complements previous work by this group examining SOX9 binding in the same samples (GSM1692996, (Ohba *et al.* 2015); as used in Chapter 3.6.1). These datasets can be compiled in the *Ccn2* enhancer UCSC Genome Browser session in order to examine the co-localisation of SOX9 and AP-1 binding and the positioning of the SOX9 sites discovered herein. According to these ChIP-Seq datasets, SOX9 and AP-1 are predicted to bind in the -137kb and -148kb enhancers which functioned in chondrocytes, in addition to the -230kb enhancer which functioned within osseous tissue (Figures 4.6- 4.8). However, in the TRAP predicted TFBS, AP-1 is only predicted to bind within -137kb. This highlights discrepancy in computationally predicted TFBS and experimentally validated ChIP-Seq interaction sites; the latter of which is more biologically relevant and reliable (Visel *et al.* 2009a).

In the -137kb enhancer, the SOX9 TFBS validated through EMSA (Figure 4.6, A) lies upstream of the AP1 site (Figure 4.6, B) and both of these sequences overlap with SOX9 ChIP-Seq sites (Figure 4.6, C). In the -148kb enhancer, ChIP-Seq peaks for AP-1 and SOX9 co-segregate into two clusters (Figure 4.7 C and D). Both of the EMSA validated SOX9 TFBS (Figure 4.7 B) lie in the 5' cluster of sites that were not present in the -148kb_short transgenic mice which had a truncated stretch of the -148kb sequence (Figure 4.7 A). This could therefore account for the reduction in chondrocyte based transgene function within the -148kb_short line compared to transgenic mice created with the whole enhancer reporter construct. A clear and conserved SOX9 consensus site within the 3' cluster of ChIP-Seq peaks was not clearly discernible in EMSA probe design, but should be investigated in future work.

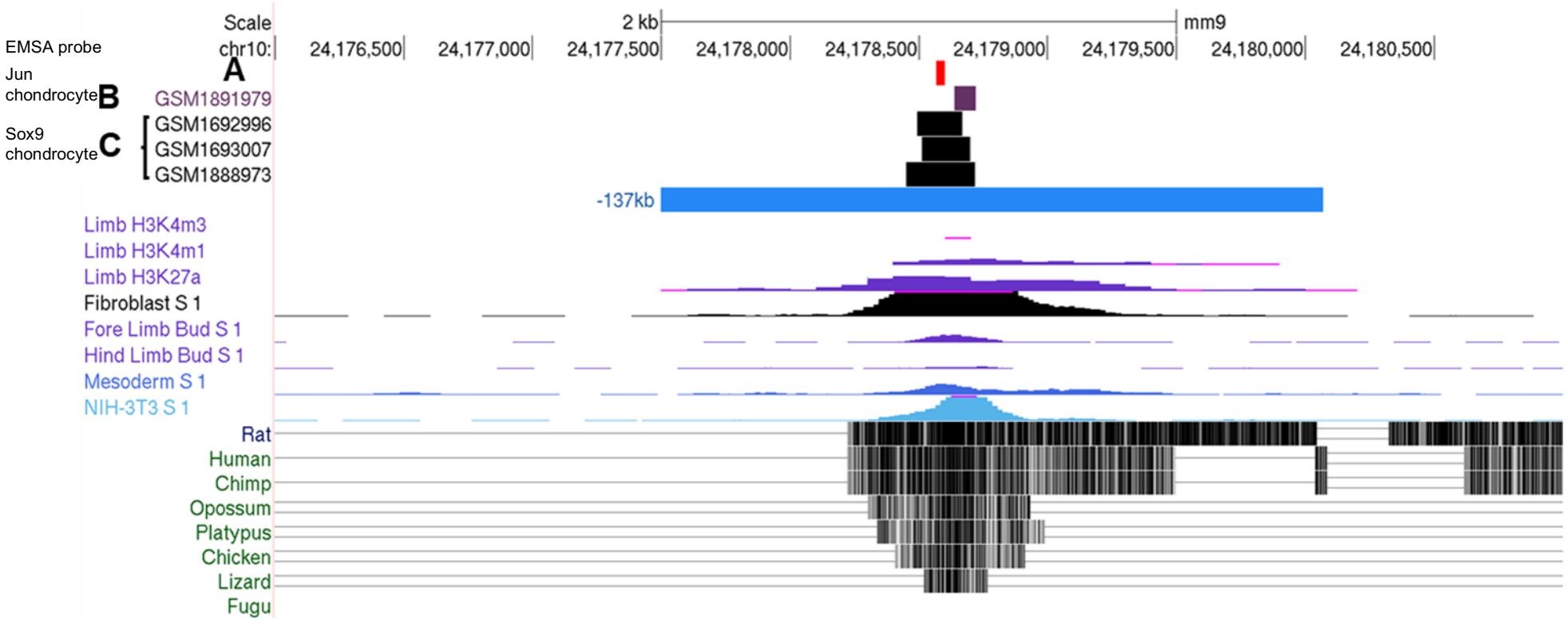


Figure 4.6: UCSC Genome Browser visualisation of SOX9 and Jun ChIP-Seq validated binding within the -137kb enhancer region. EMSA validated SOX9 binding motif (A) lay within close proximity to ChIP-Seq peaks for Jun TFBS (B, GSM1891979) and SOX9 (C,) (Garside *et al.* 2015; Ohba *et al.* 2015), and in close proximity to a Jun ChIP-Seq peak (B) (He *et al.* 2016). In line with findings of He *et al.* (2016), Jun may interact with SOX9 in order to drive the function of the enhancer in chondrocyte specific expression of *Ccn2*

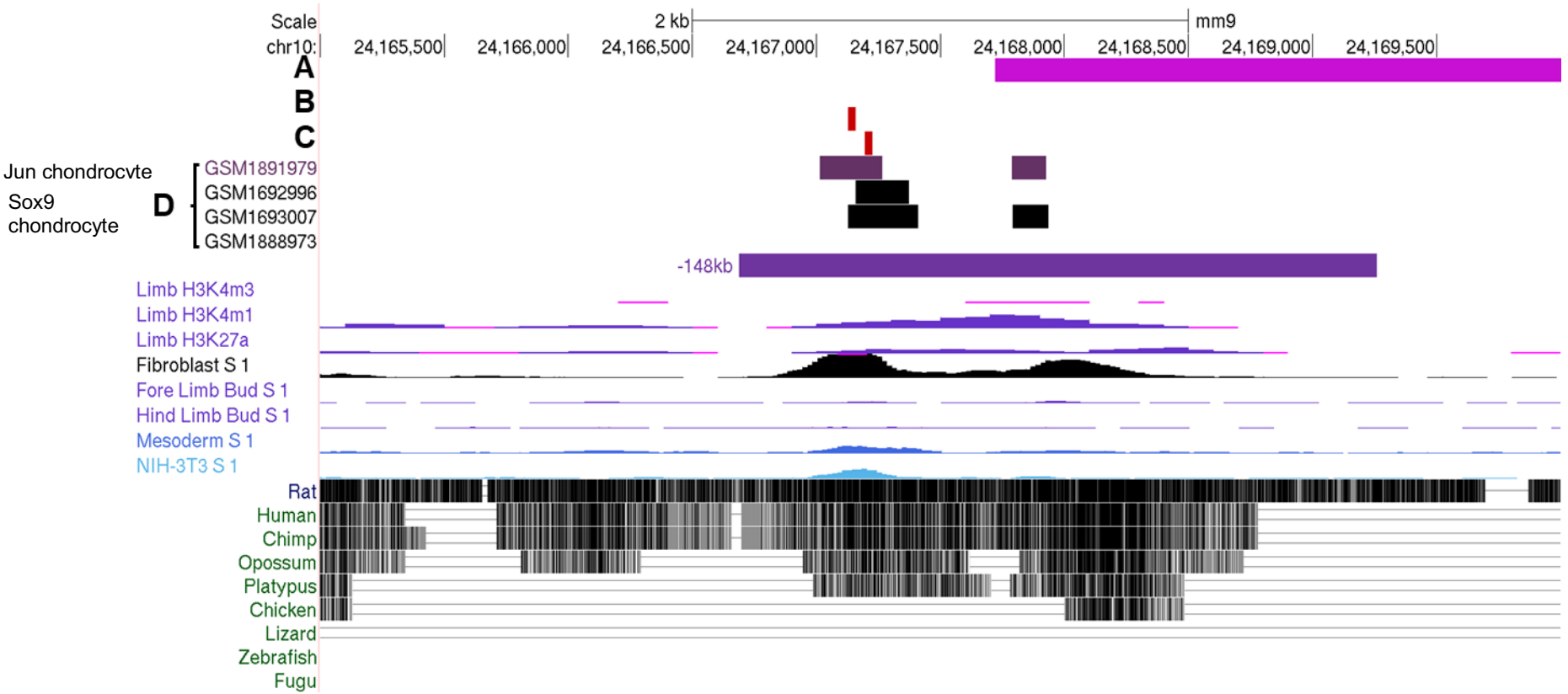


Figure 4.7: UCSC Genome Browser visualisation of SOX9 and Jun ChIP-Seq validated binding within the –148kb enhancer region. EMSA validated SOX9 binding motifs (B and C) lay upstream of the -148kb sequence used for the -148kb_short transgenic mouse line (A). ChIP-Seq peaks for chondrocytic SOX9 and Jun interaction were split within two clusters, as with other enhancer associated attributes and sequence conservation.(He *et al.* 2016; Ohba *et al.* 2015) The EMSA validated SOX9 TFBS lay within both SOX9 and Jun peaks of interaction (D), supporting the study of He et al. (2016) with the notion that both of these TF direct chondrocyte transcription of *Ccn2* through this enhancer element.

In accordance with the findings of He *et al*. (2016), the presence of both transcription binding sites would suggest that both AP-1 and SOX9 act synergistically in facilitating the activity of the enhancer and transcriptional output from *Ccn2*. Figures 4.6 and 4.7 therefore would suggest that Jun and SOX9 co-localise within the -137kb and -148kb enhancers with concurrent function of both TF increasing the transcriptional output from *Ccn2* via enhancer regions. These samples were isolated from postnatal tissue suggesting that the -137kb enhancer functions later than at the E15.5 time-point examined with the current *in vivo* reporter assay system. Further pertinent experiments would involve ratification of this theory through assay of enhancer function at P1 and comparison with E15.5. In addition, ChIP-Seq experiment using articular chondrocytes isolated from E15.5 embryos would enable elucidation as to whether interaction with both AP-1/Jun and SOX9 combined boosts enhancer activity.

The -230kb enhancer also contains sites for both AP-1 and SOX9 (Figure 4.8), but *in vivo* reporter assay suggests that this region is more active in osteoblastic cells than chondrocytes. On the basis of the findings of He *et al*. (2016) it would therefore be expected that the enhancer most probably interacts with one of the transcription factors, and that the other blocks the function of the enhancer within chondrocyte cells. Of course, this data doesn't account for protein-protein interaction which may be the manner in which chondrocyte based enhancer activity is blocked.

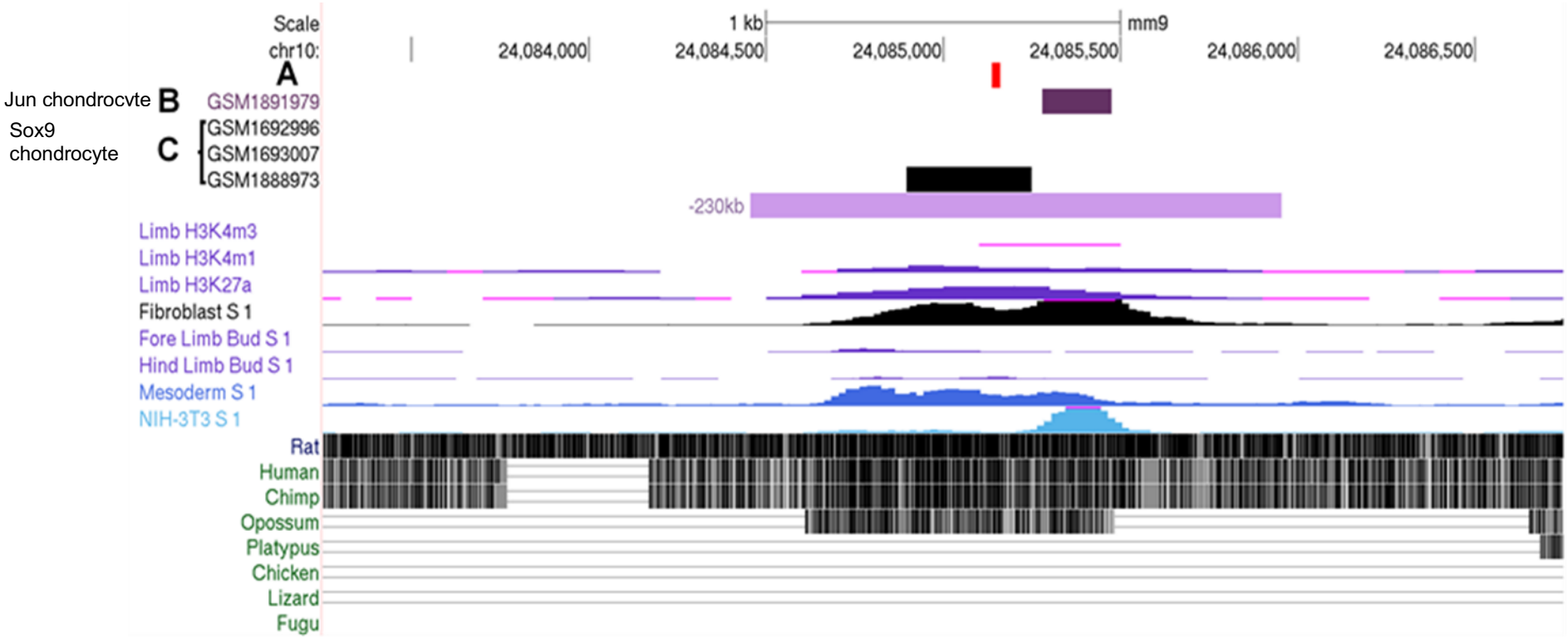


Figure 4.8: UCSC Genome Browser visualisation of SOX9 and Jun ChIP-Seq validated binding within the –230kb enhancer region. EMSA validated SOX9 binding motif (A) lies within a SOX9 ChIP-Seq peak (C,(Garside *et al.* 2015)) but not within Jun ChIP-Seq peak. The lack of transgene activity in chondrocytes as a results of -230lkb function suggests that the TF may not bind cooperatively, and may antagonise one another in order to prevent chondrocyte based enhancer function, in line with the findings of He *et al*. (2016).

At E15.5, for founder A of the -230kbHsp68LacZ line, the enhancer was active within periosteal cells, and osteoblasts within long bone primary ossification centre. There was no transgene activity for E15.5 derived from founder C, but at P7 and six weeks of age the enhancer functioned within osseous tissue. CCN2 has been proven to function in both chondrocyte and osteoblastic cells (Arnott *et al.* 2011; Ivkovic *et al.* 2003). If AP-1 and SOX9 bind in tandem to cognate DNA motifs within -230kb to drive enhancer function in line with the findings of He *et al*. (2016), the enhancer may function in trans-differentiation of hypertrophic chondrocytes to osteoblasts, with promotion of osteoblastic behaviour. This theory could be tested using a Cre-recombinase approach. A transgenic mouse construct containing *Cre* under control of the enhancer region would lead to Cre expression in cells where the enhancer was active. Crossing this mouse line with a reporter transgenic line such ROSA$^{mT/mG}$ would enable tdTomato (mT) signal to be observed in cells where the enhancer was inactive. Where the enhancer was active and Cre was produced, in the presence of tamoxifen, cells in which the enhancer was active and resultant daughter cells would produce GFP (mG) signal. Therefore, for -230kb, if the enhancer was inactive at E13.5 in chondrocytes signal red mT would be abundant. By E15.5, if the enhancer was active in hypertrophic cells green mG signal would be observed. In osteoblastic cells derived from these chondrocytes, mG signal would also be produced at later time-points. This would therefore allow lineage tracing of enhancer activity. Moreover, the role of CCN2 in chondro-osseous lineage switching could also be explored; a concept which has not been elucidated. This lineage-tracing approach could be applied for any of the enhancers described herein, in order to ascertain the role of the enhancers in cell lineage fate commitment. This a highly relevant question given the importance of enhancer function in cellular programming (Bogdanović *et al.* 2012; Huang *et al.* 2016)

A further consideration in the osteoblastic expression observed for the -230kb enhancer is the role of Wnt signalling. This pathway has been well documented as a repressor of activity (Kozhemyakina *et al.* 2015). Wnt signalling through β-catenin and SOX9 have been proven to antagonise the function of one another in skeletal development (Akiyama *et al.* 2004). Both SOX9 and TCF-LEF contain high mobility group box domains which mediate binding of the protein similar target consensus sequences DNA target resulting in DNA bending (Karsenty and Wagner 2002). The work of Huang *et al.* (2010) sets a precedent for antagonism between SOX9 and TCF-LEF/β-catenin in the regulation of *Ccn2*, in the vicinity of the promoter, as aforementioned in Chapter 1.4. Therefore, TCF-LEF/β-catenin may interact with

the -137kb and -230kb enhancers repressing SOX9 activity and chondrocyte based *Ccn2* expression. This change could occur in late endochondral ossification with the switch from the dominance of cartilage to osseous tissue (Mackie *et al.* 2008). Within the EMSA probes designed for the putative SOX9 binding sites for -137kb and -230kb, TCF-LEF was also predicted to bind. Whilst beyond the scope of the current project, it would be pertinent to examine in further EMSA experiments whether TCF-LEF interacts with the enhancers through these sequences.

4.2.2 Do other transcription factors regulate chondrocyte based -148kb function?

Given the high levels of gene transcription driven by the -148kb enhancer within chondrocytes, it would be assumed that the master regulator of chondrocytes, Sox9 would be the lynchpin TF in the function of this enhancer region. This regulation was not immediately clear with a lack of SOX9 consensus sites using the TRAP tool (Chapter 5.5). Nonetheless, SOX9 was proven to bind to sequences within this enhancer region. However, these EMSA validated TFBS (Chapter 3.8.2) lay within the 5' sub-region of -148kb that was not used to create the -148kb_short Hsp68LacZ transgenic mice, yet chondrocyte expression, albeit weak, was still observed within founder variants of this line. This therefore piques interest as to whether SOX9 is the lynchpin transcription factor in -148kb chondrocyte function. Despite the strong consensus as to SOX9 being the master regulator of chondrocytes, its role as a pioneer factor in chondrocytes has recently been questioned. Liu *et al*. (2018) demonstrated that during chondrogenesis, SOX9 is not required for the removal of repressive epigenetic state and chromatin reprogramming; therefore SOX9 did not behave as a pioneer factor. This suggests that further transcription factors are required as pioneer factors in the commissioning of -148kb enhancer function in order to drive gene transcription during chondrogenesis.

The results from EMSA analysis of RUNX2 TFBS within -148kb (Chapter 3.8.3) did not conclusively prove TF-enhancer sequence interaction. Of the three probes examined, -148kb RUNX2_3 is more likely to be bound by RUNX2 with the small shift in probe signal indicating that there may be weak interaction between the protein and DNA. In addition this interaction was reduced with the use of wild-type competitor oligonucleotide in a dose dependent manner.

The fact that the enhancer functioned in the articular cartilage of both CTGF148 and -148kb_short mice suggests that a factor within the 1.5kb enhancer fragment mediates chondrocytic function. The markedly stronger transgene expression within articular cartilage of CTGF148 mice compared to -148kb short could have been due to chondroprotective SOX9 interaction in the 5' sub-region of the enhancer that was not present in the short transgene construct (Henry *et al.* 2012). The β-galactosidase expression driven by -148kb was observed in articular cartilage that had physiological appearance rather than with osteoarthritic degradation or clefts, suggesting that -148kb is not regulated by RUNX2 within adult articular cartilage. RUNX2 in articular cartilage encourages hypertrophic phenotype and induction of the expression of catabolic genes such as *MMP13*, and therefore OA like changes in cartilage matrix (van der Kraan and van den Berg 2012). Loss of enhancer function over time tallies with chondroprotective role for CCN2 with enhancer loss of activity and therefore CCN2 within articular cartilage with age. RUNX1 could also regulate the enhancer through interaction with consensus sequences that are redundant with those of Runx2. Furthermore, RUNX1 expression has been observed in non-OA articular cartilage, with a suggestion that this transcription factor is involved in superficial zone chondrocyte proliferation (LeBlanc *et al.* 2015). RUNX2 could regulate enhancer function within hypertrophic chondrocytes of the growth plate which also demonstrated *in vivo* transgene activity in adulthood. These findings highlights the issue of trying to predict TFBS based on ChIP-Seq data from other tissues, as with the use of RUNX2 ChIP-Seq in trying to predict RUNX2 sites within chondrocytes. ChIP-Seq interrogation of the interaction between enhancer sequences and RUNX1/2 in various populations of chondrocytes at numerous time-points would yield powerful information with regards to the capacity of Runx to regulate enhancer activity.

ChIP-Seq data also suggests that DLX5 is capable of binding to sequences within -148kb during osteoblastic differentiation (GSE76185, Hojo *et al*. (2016)), as illustrated in Figure 4.9. DLX5 has been proven to be involved in the promotion of chondrocyte maturation and transition to chondrocyte hypertrophy in endochondral ossification (Ferrari and Kosher 2002; Zhu and Bendall 2009). Therefore, it is logical that the ChIP-Seq peaks contain TFBS that are utilised by DLX5 in mediating enhancer function in driving *Ccn2* transcription in the promotion of chondrocyte hypertrophy. Further characterisation of DLX5 regulatory sequences in -148kb through EMSA is required in order to validate this idea.

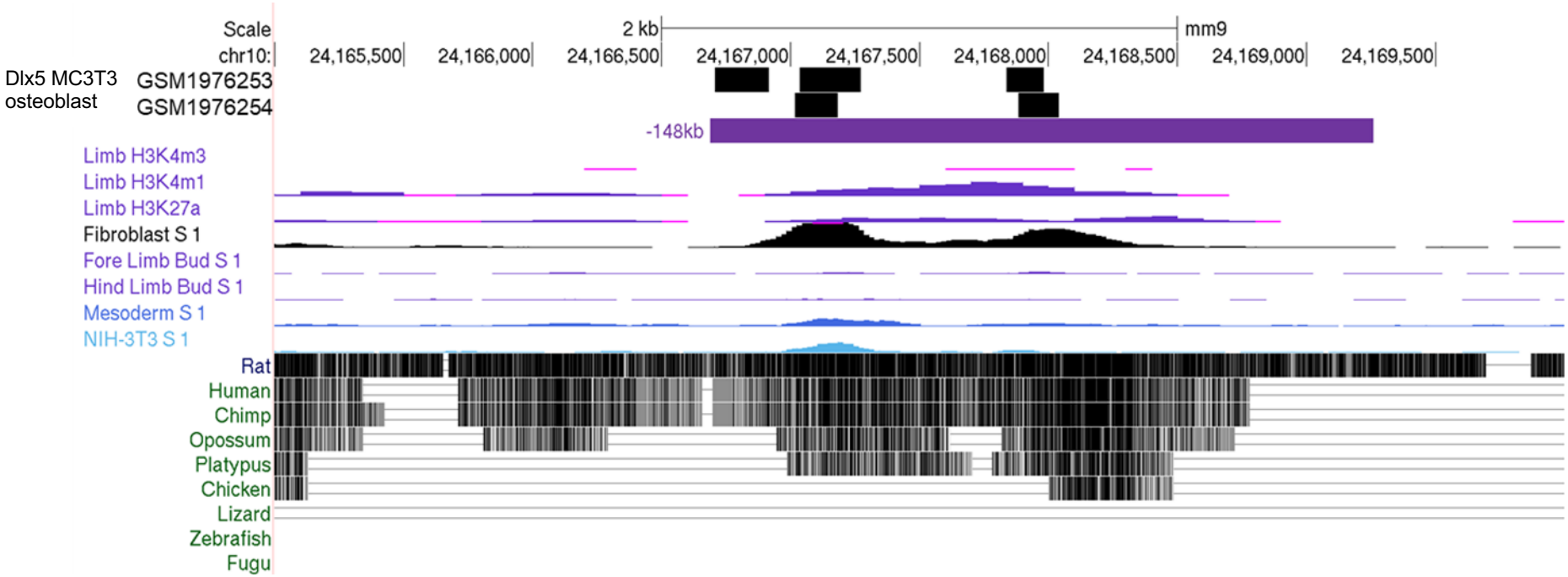


Figure 4.9: UCSC Genome Browser visualisation of publicly available ChIP-Seq data for Dlx5 binding within -148kb. ChIP-Seq peaks for interaction between DLX5 (Hojo *et al.* 2016) and the -148kb enhancer in MC3T3-E1 pre-osteoblast cell line are grouped in two clusters, akin to those for SOX9 and Jun (Figure 4.7) and as was observed with enhancer associated chromatin characteristics in addition to sequence conservation which formed the rationale for the creation of the -148kb _short line (Chapter 3.5). This suggests DLX5 could bind within both the -148kb full and -148kb short sequences.

Upon examination of the -148kb sequence using the TRAP TFBS prediction tool (Chapter 5.5), several homeobox (HOX) transcription factors are highly ranked as interacting with the enhancer region. These paralogous genes are well established as regulators of skeletogenesis (De Laat and Duboule 2013). For example, removal of *HOX10* and *HOX11* gene groups each causes dysmorphism in vertebral, limb and synovial joint development (Koyama *et al.* 2010; Wellik and Capecchi 2003). Several HOX10 and HOX11 TFBS are predicted within -148kb, suggesting that these transcription factors may dictate enhancer function and therefore *Ccn2* expression in skeletogenesis. In addition HOX11 has been shown to regulate endochondral ossification from the perichondrium, thereby concurring with endogenous *Ccn2* expression (Ivkovic *et al.* 2003; Papenbrock *et al.* 2000).

Gross *et al*. (2012) demonstrated the role of *HOX11* genes in chondrocyte differentiation, with knockout of *HOXA11* and *HOXD11* led to repression of columnar chondrocyte morphology, chondrocyte hypertrophy and the expression of *Runx2*. Application of this finding to the current work suggests that there may be interaction between HOX11 transcription factors and RUNX2 at the -148kb enhancer, mediating continued stringent control of *Ccn2* transcription in the successive differentiation of chondrocytes along the endochondral growth plate. There could also be redundancy in the function of HOX genes in the regulation of -148kb, which may in part account for the fact that several TFBS are predicted in the sequence. This would ensure enhancer function in the event of mutation of a TFBS and therefore lends credence to notion that the enhancer is important in driving *Ccn2* expression during development and skeletogenesis. The prevalence of HOX genes and the sequential and tightly controlled activation in embryogenesis, as opposed to function within postnatal tissue, would suggest that these TFBS are not used to drive the postnatal function of the enhancer observed in CTGF148 and founder D of -148kb_short transgenic animals. Further *in vitro* validation of this idea through the use EMSA or ChIP-Seq is required.

Interaction with additional limb related transcription factors could also refine enhancer function with chondrocyte specificity in function. PITX1 is a transcription factor involved in chondrogenesis and skeletal development (Lanctôt *et al.* 1999). As of yet, PITX1 has not been associated with the regulation of *Ccn2* expression. However, using the TRAP tool, there are predicted TFBS for PITX1 in both full and short variants of the -148kb enhancer sequence. Beyond the scope of the current

project, PITX1 should be examined as a potential regulator of both -148kb enhancer and *Ccn2.*

4.2.3 How stringent is enhancer grammar in function?

Regardless of the identity of the transcription factors that bind to the enhancers of *Ccn2*, an important aspect of elucidating enhancer function will be characterisation of the hierarchical protein interactions that facilitate chromatin topological organisation and ultimately interaction between enhancer and *Ccn2* promoter region.

The fact that a shorter version of the -148kb enhancer containing the approximately 1.5kb of sequence, rather than the full region of approximately 2.5kb, was able to drive gene transcription would suggest that the enhancer doesn't follow the enhanceosome model of enhancer grammar. As outlined in Chapter 1.3.3, enhanceosome function relies on stringent interaction of transcription factors with an enhancer sequence in a highly ordered manner (Spitz and Furlong 2012). The loss of 1kb from the -148kb sequence did not impinge on the capacity of the region to drive transgene expression. The 5' 1kb fragment of the enhancer therefore must contain TFBS which are dispensable, and may be redundant in enhancer function. Moreover, redundancy in TFBS positioning may favour a more robust billboard model-like function whereby the presence of the TFBS, rather than positioning within the sequence is required for enhancer function (Kulkarni and Arnosti 2003). The chondrocyte based transgene expression in founders bearing the -148kb_short construct suggest that chondrocyte related TFBS are found within this shorter sequence. However, the variation in the tissues in which transgene activity was observed in -148kb_short would suggest that regulatory interactions constraining cell lineage-specific enhancer function were lost. Therefore, the fidelity of enhancer function is bestowed by the full sequence which reduces gene transcription within non-chondrocyte cell types, which could occur through repressive TF interaction at the enhancer. The expression within the -148kb_short transgenic animals was predominantly within mesenchymal tissues, suggesting that the factors that bind to this region in order to modulate its function are mesenchymal cell lineage related, but not cell-type specific. The enhancer could also follow a TF collective model of function whereby protein-protein interaction allows enhancer function rather than TF-DNA interaction (Harmston and Lenhard 2013). Beyond the scope of the current work in terms of time and budget, further

characterisation of the TFBS within -148kb and their utilisation is required in order to understand the signalling mechanisms and interactions that control enhancer activity. This also applies for each of the enhancers described herein, in order to understand the grammar that underpins both enhancer DNA-TF interaction and resultant enhancer-promoter interaction.

The stringency of enhancer grammar is also raised with the observation that in the CTGF148 -148kb full line, and -148kb_short transgenic mice, the cell lineages in which enhancer function was observed changed over time. This contradicts the consensus that enhancers function in lineage-specific regulation of gene expression (Heinz *et al.* 2015; Li *et al.* 2016). For the CTGF148 line, between E11.5 and E13.5 expression within the vasculature and heart was observed before a switch to only chondrocyte based reporter gene expression in later embryonic development and adulthood. For the -148kb_short, expression in progeny of Founder C was observed in perichondrium/ primitive periosteal tissue at E15.5, yet at potently functioned within the musculature at approximately six months of age. The progeny of founder D did not exhibit vasculature localisation of transgene expression, which was potent by six months of age. This effect could be due to the site of integration where transgenic constructs landed in the genome, with influence from other regulatory sequences within the vicinity reducing the specificity of *LacZ* expression to enhancer activity. Further analysis of enhancer function *in vivo* is required to understand this.

4.3 Further Ccn2 enhancer function

4.3.1 Are there further enhancers of *Ccn2*?

Reporter gene expression driven by the enhancer regions described herein does not constitute the replete expression profile of *Ccn2.* For example, the -230kb, -198kb, -148kb, -137kb and -102kb enhancers did not drive expression in the gastrointestinal system, neuronal tissue, dentition and reproductive system (Friedrichsen *et al.* 2003, 2005). This therefore suggests that further *cis*-acting regulatory elements are required in the regulation of *Ccn2* transcription. As of yet the only studies examining the region downstream of *Ccn2 in vivo* has been those of Hall-Glenn and Lyons (2011) and Hall-Glenn *et al*. (2012) who utilised the previously mentioned Tg(Ctgf-EGFP)FX156Gsat mouse line. Neither of these studies attempted to define enhancer elements within the 160kb of sequence used

in this BAC transgenic construct. The locus encompassing this can be examined in the UCSC Genome Browser in order to identify further putative enhancers of *Ccn2* expression (Figure 4.10). Using the same approach as used in the current project, within the UCSC Genome Browser, there are some potential regulatory elements further downstream of *Cc2* however, there are no clear enhancer regions with similar peaks for H4K4me1 and H3K27ac comparable to the enhancers described in the current study.

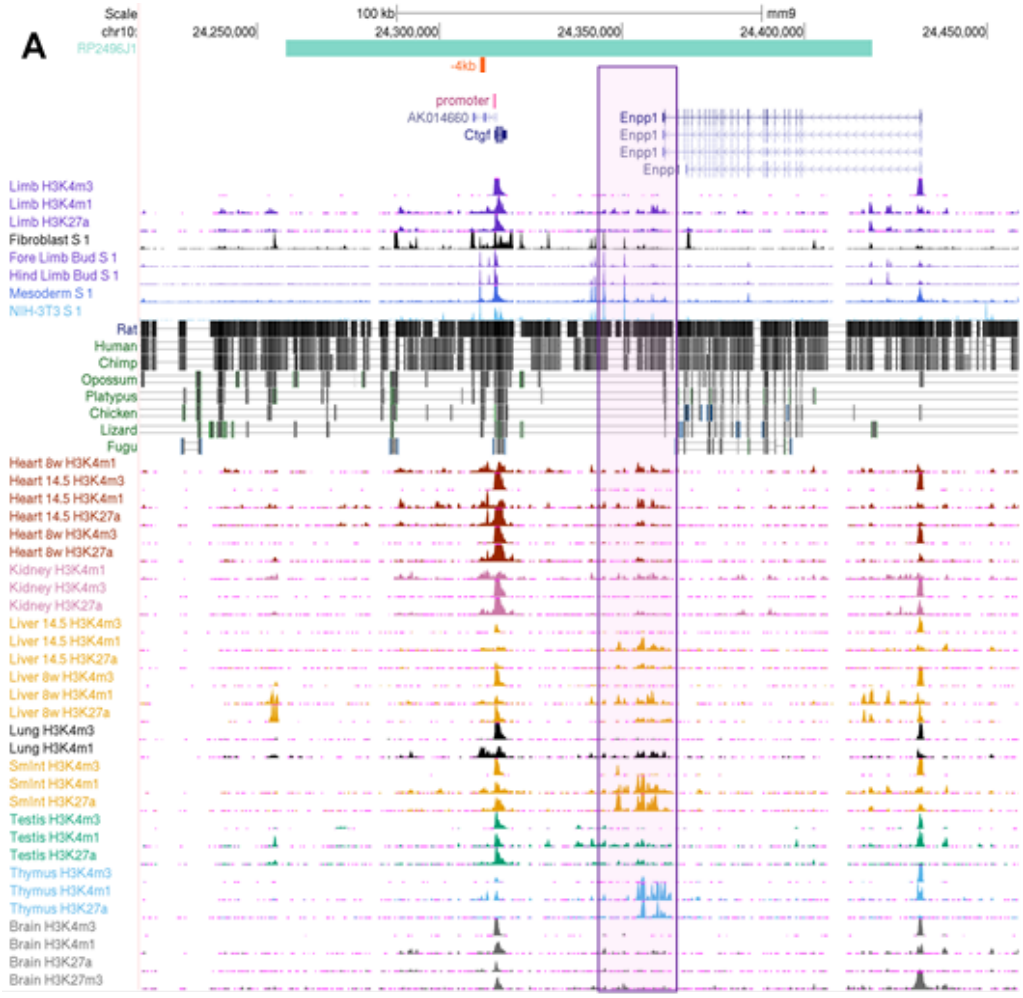


Figure 4.10: UCSC Genome Browser graphic of regulatory element-associated histone modification in several tissues types in the loci of interest the Tg (Ctgf-EGFP) FX156Gsat. Within the BAC construct (A, green bar), the most significant peaks across tissues occur in the promoter region for *Ccn2*. There are some small peaks for H3K4me1 and H3K27ac in a cluster of loci (purple box) approximately 35kb downstream of *CCN2* in several tissues, however this region could regulate the expression of *Enpp1* rather than *Ccn2*.

The lack of clearly discernible regulatory elements downstream of *Ccn2* suggests that there are additional non-coding regulatory elements upstream of *Ccn2.* The current project was concerned primarily with skeletal; and especially chondrocyte, based enhancer function, it was therefore not feasible to examine each tissue type.

It is therefore plausible that further enhancers regulate the expression in *Ccn2* in other tissue types. Using the approach of the current project, tracks concerning multiple tissues can be used in the identification of further *Ccn2* enhancers located in the 300kb upstream intergenic region (Figures 4.11 and 4.12).

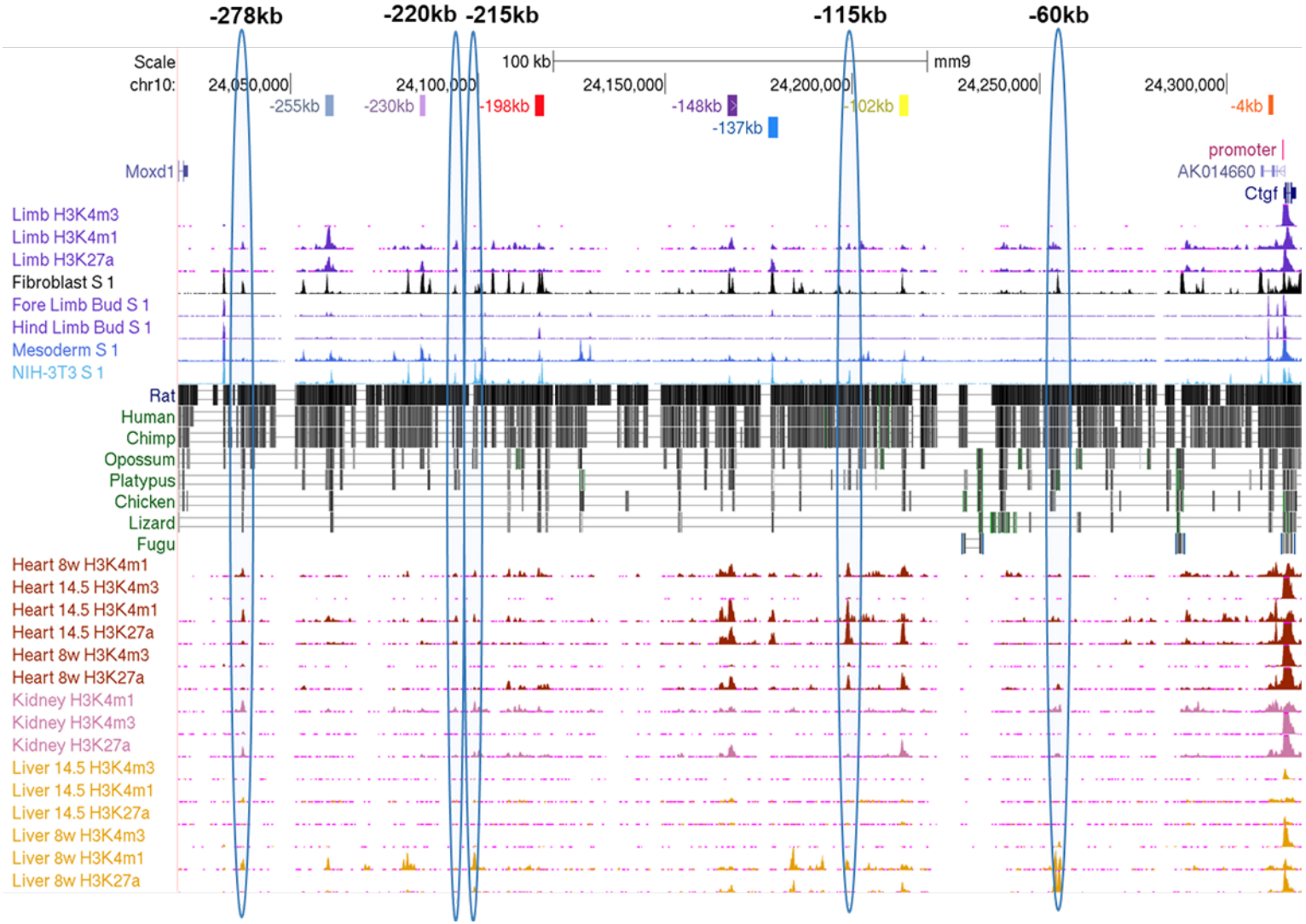


Figure 4.11: Use of the UCSC Genome Browser to predict enhancers of *Ccn2* active in heart, kidney and liver tissue. Three further enhancers are discernible located -278kb, -215kb -115kb and -60kb upstream of *Ccn2*. There are peaks for sequence conservation and enhancer associated chromatin in heart, kidney and liver samples for the -278kb and -215kb regions. The peaks for H3K4me1 and H3K27ac occur most prominently in -115kb for heart samples, suggesting an enhancer that functions specifically within heart tissue at E14.5, and to a lesser extent at 8 weeks of age. Peaks for enhancer associated chromatin characteristics are also found within the enhancers examined herein.

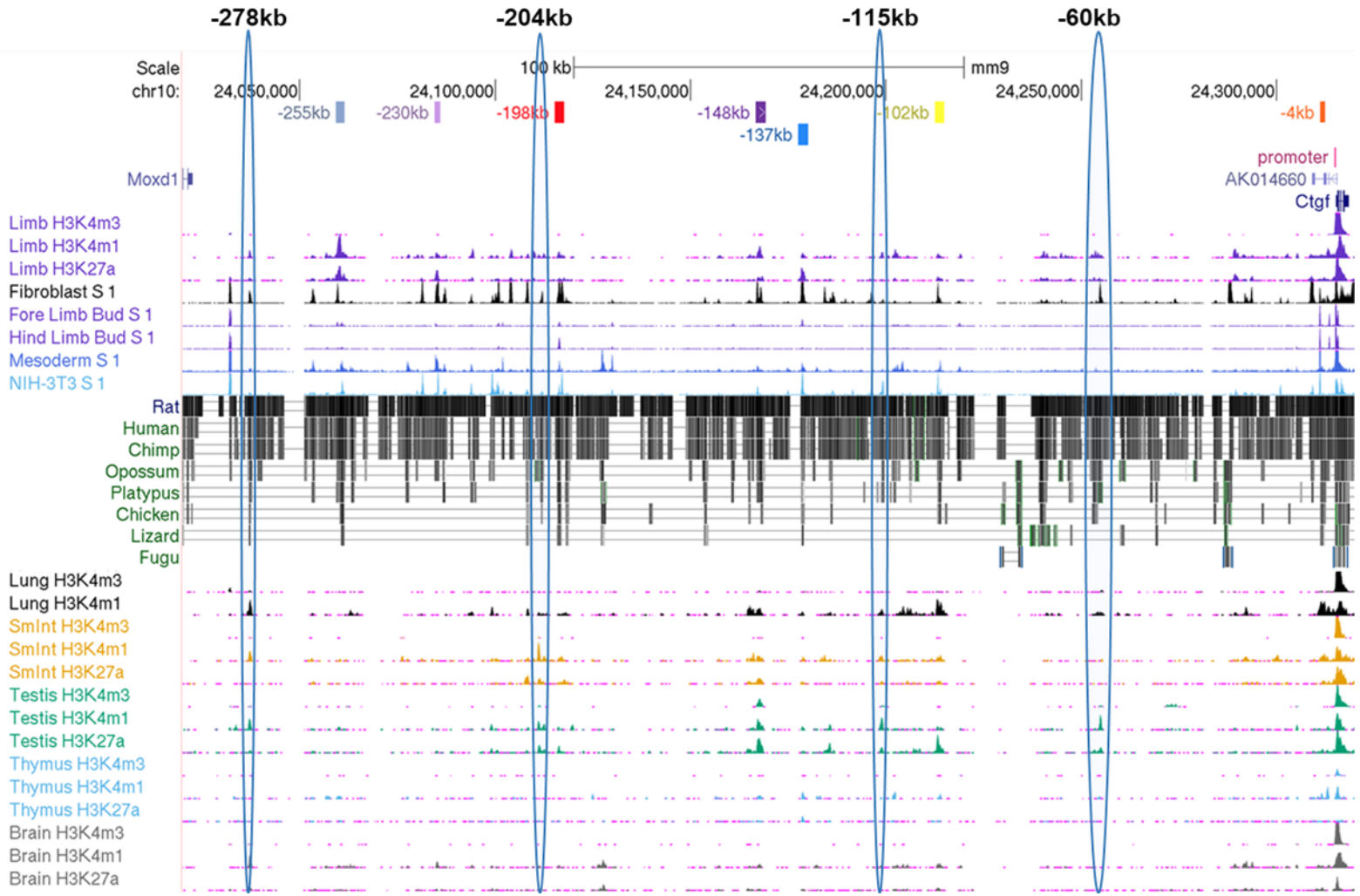


Figure 4.12: Use of the UCSC Genome Browser to predict enhancers of *Ccn2* active in lung, small intestine, testis, thymus and brain tissues. In addition to the -278kb, and -115kb candidate enhancers identified in Figure 4.11, further enhancers are predicted at -204kb in small intestinal and testis tissues, and at -60kb in testis tissue. The pre-existing putative enhancer regions also contain peaks for H3K4me1 and H3K27ac in several tissues.

Validation of the new putative -278kb, -220kb -215kb, -204kb, -115kb and -60kb enhancers using the transgenic approach utilised in the current study would be vital in understanding further temporospatial determinants of *Ccn2* transcription. The peaks within a single enhancer across several tissues indicate that enhancer function is more likely to be cell lineage-specific rather than tissue specific, as there may be several cell types present in a tissue, yet the intracellular signalling mechanisms that underpin transcriptional regulation may be similar between the cell type of common lineage across several tissues, rather than the many cell types of one tissue. For example, the full -148kb enhancer region contains peaks for both H3K4me1 and H3K27ac; denoting active enhancer activity, within the heart at E14.5 and 8 weeks, small intestine and testis. This theory could be validated using a cell-specific rather than tissue based ChIP-Seq approach. Although, increased accessibility and activity -148kb chromatin in heart tissue could account for why reporter gene expression was observed in the heart in the CTGF148 line during embryonic development and in mice harbouring the shorter version of this enhancer in adulthood. Validation of the putative -220kb, -115kb and -60kb would be important in the context of the current project as they each contain peaks within limb based tissue, whilst overlooked thus far, they too could regulate chondrocyte based expression of *Ccn2*.

4.3.2 Do enhancers of *Ccn2* transcription function at further time-points, or in disease?

Whilst beyond the capabilities of the current project, examination of activity of each enhancer at several time-points would enable greater understanding of the regulatory cues that govern enhancer function in the temporospatial regulation of *Ccn2* transcription during development and adulthood. It would also be important to examine the capacity of the enhancers found here to function in adulthood. This was achieved for the CTGF148 line which demonstrated that the enhancer functioned in chondrocytes from embryonic chondrogenesis and postnatal articular cartilage. At five months of age, the -148kb enhancer still functioned in cartilaginous elements; albeit to a lesser extent, with reduced transgene activity compared to earlier time-points. This suggests that -148kb is required in the both the establishment and maintenance of *Ccn2* expression within cartilage. The enhancer could carry on driving transgene expression at a later time point such as 12 months of age, which was not feasible to assay in the current study.

A pertinent question is whether the enhancer is used in solely a housekeeping-like fashion, or whether the output from this enhancer can be increased in a latent manner (Ostuni *et al.* 2013; Zabidi *et al.* 2015). This could be tested utilising the CTGF148 line and mechanical loading model of osteoarthritis (Poulet *et al.* 2011). This is relevant for the -137kb enhancer too, given the articular localisation of enhancer function. The knees of transgenic mice could be loaded in order to induce OA changes in the articular cartilage and synovial joint. Comparison of X-gal staining intensity between loaded and contralateral non-loaded knee joints would allow changes in enhancer activation to be observed. Given the controversial role of *CCN2* in osteoarthritis, it would be interesting to examine whether enhancer activity was increased to boost the amount of CCN2 in a chondroprotective manner, or enhancer activity would be repressed to prevent the accumulation of CCN2 within the joint and subsequent catabolic degradation of cartilage matrix. Therefore, the activation of the enhancer may change with the severity of the disease. TNF-α and TGF-β have both been implicated in osteoarthritis (Gabay *et al.* 2012; van der Kraan and van den Berg 2012), and both factors have previously been shown to antagonise the function of one another at the *CCN2* promoter (Abraham *et al.* 2000), this interaction could be repeated in the enhancer elements. In addition, TNF-α has been demonstrated to induce *CCN2* expression in rheumatoid arthritis, leading to osteoclastogenesis (Nozawa *et al.* 2009). Further work in order to examine the TFBS and mechanism for osteoarthritis-linked signal transduction is required in order to ascertain the validity of this theory. Orthologous sequences of the *Ccn2* enhancers could also be examined in patients with osteoarthritis. As outlined in Chapter 1.3, SNP within enhancers are increasingly being examined in disease association studies (Andersson *et al.* 2014). Capellini *et al*. (2017) found that an SNP in an enhancer for *GDF5* causes reduced gene expression in the epiphyseal region of long bones, associated with OA susceptibility postulating that this increases susceptibility to osteoarthritis. This echoes the work of Reynard *et al*. (2016) who found that SNP within an intronic enhancer of cartilage related gene *CHST11* cause changes in protein affinity for TFBS with decrease in enhancer activity and therefore target gene expression leading to OA susceptibility. This sets a precedent that SNP within the enhancers of *CCN2* could lead to increased susceptibility to OA. Furthermore, predisposition to other diseases caused by changes in *CCN2* expression such as fibrosis could be caused by SNP within the enhancers described herein.

The expression of *CCN2* has previously been shown to be induced in proliferating periosteal cells in bone fracture repair (Nakata *et al.* 2002). This therefore leads to the question as to whether the -230kb enhancer is activated in a latent manner in order to drive osteoblast based CCN2 production in fracture healing. The function of further enhancers could also be latently induced in the maintenance of tissue homeostasis. For example, Kapoor *et al*. (2008) found activation of the promoter region in a skin wounding wounded model. This could occur in conjunction with latent enhancer activation that drives gene expression as part of a concerted response to increase ECM components and promote repair.

4.4 Future direction

In summary, the findings described herein give an insight into the role of novel *cis*-acting regulatory elements in the region upstream of *Ccn2*. However, further work in required in order to fully comprehend the narrow temporospatial frames in which enhancers function to control the transcription of *Ccn2*, and the complicated molecular regulatory interactions that facilitate this process. Given that enhancers function within specific temporospatial windows, it is important to have a global approach in study of their function. This is more readily accomplished through the use of *in vivo* methodology.

An important aspect of future should be study of the human *CCN2* locus and validation of function between the murine enhancer regulatory sequences described herein and human regulatory elements. The congruity of predicted TAD boundaries between human and mouse shown in the 3DIV and Dixon *et al.* (2012) datasets would suggest that enhancers of *CCN2* would be found in the equivalent positions to those described herein. Enhancer sequences were visualised in the human genome in the current project, but this was mostly carried out for contextualisation of TFBS. Therefore, future would should assay the ability of orthologous human *CCN2* enhancer sequence to drive reporter gene expression, in addition to characterisation of loci through the wide array of *in silico* resources available such as SCREEN, the ROADMAP project and ChIP-Seq datasets from human samples. Use of the Genehancer dataset available as part of the 'Genecards' website (https://www.genecards.org/) will also be a key aspect of this as it contains association data from interaction (Hi-C) and eRNA expression for predicted enhancers of *CCN2*, of which there are several entries. These regions

overlap with several of the murine enhancers described herein and are therefore fundamental in understanding *CCN2* transcriptional regulation.

A priority in future research into *Ccn2* enhancers should be examination of the topological interactions that underpin the endogenous utilisation of enhancer regions in orchestrating the transcriptional regulation of *Ccn2*. This could be carried out using high-throughput chromatin conformation capture based assays, for example high resolution Hi-C in order to visualise interactions between enhancers and the promoter region, in addition to verification of TAD structure (Whalen *et al.* 2016). In using samples from various cell and tissue types, the interactions within different cellular contexts would be visualised and compared. This would enable greater understanding of the temporospatial specificity in *Ccn2* enhancer function and the mechanisms that dictate this key attribute.

A critical aspect of future work in characterising the enhancer regions described herein will be the manipulation of the enhancer sequences *in vivo* through CRISPR-Cas9 technology. Whilst unsuccessful in the application to make enhancer knockout transgenic mice in the current project, the results from this process will nevertheless inform and aid future work. Despite RNA-Cas9 injection into the cytoplasm being posited as the most efficient CRISPR-Cas9 system (Horii *et al.* 2014), this was not effective under the condition used herein. The strategy employed in order to identify and design guide RNA was successful; as shown by the *in vitro* cleavage of target DNA substrate (Chapter 3.8). The lack of mutant mice highlight that the process required further optimisation at the microinjection stage. For example, although the recombinant Cas9 protein used had a nuclear localisation signal, injection of complexes directly into the pronucleus rather than cytoplasm would boost increase the amount of CRISPR-Cas9 complexes within the nucleus and therefore genomic targeting; albeit with higher rates of mortality.

CRISPR-Cas9 has been well established in the study of enhancers and optimisation of the technique used in the current study would herald a turning point in the ability to assay the endogenous function of *Ccn2* enhancers (Catarino and Stark 2018; Fulco *et al.* 2016; Lopes *et al.* 2016). Ideally, this work would also use *CCN2* reporter mouse whereby endogenous expression of the gene would be replaced with that of a reporter such as GFP; which could be achieved through CRISPR-Cas9 genome editing. This would allow manipulation of enhancer function to be assayed through the production of GFP. This would also be the most effective

technique in maintaining the endogenous chromatin topology and environment that enhancer function depends on. Furthermore, this style of reporter gene assay would negate the issues site integration and construct copy number that could have hampered the accuracy of the current study.

The first step using this approach would be deletion of enhancers, as was attempted herein. This would enable the importance of the function of each enhancer in physiological *Ccn2* expression to be examined. For example, if the -148kb enhancer is the lynchpin regulator of chondrocyte based *Ccn2* expression, excision of this enhancer would be expected to lead to recapitulation of the profound chondrodysplasia phenotype exhibited by *Ccn2*$^{-/-}$ mice (Ivkovic *et al.* 2003). Transgenic mice lines with deletion of each enhancer could be crossed in order to determine how robust the actions of *cis*-acting regulatory elements are in the regulation of *Ccn2*. This approach could also aid in determination of redundancy in enhancer function and whether any of the regions function as shadow enhancers, and their function is dispensable in the physiological expression of *CCN2*. Moreover, CRISPR-Cas9 has previously been used to demonstrate functional hierarchy of enhancers (Huang *et al.* 2016). If this were applied to the current project, would -148kb be found to be the most important enhancer of *Ccn2*?

The robustness of enhancer function could also be tested in CRISPR-Cas9 based removal of sub-regions of enhancers. This reductionist approach would enable identification of the key sequences within the enhancers that are responsible for regulation and function. In addition, individual TFBS could be deleted or mutated using CRISPR-Cas9. Furthermore, enhancer grammar could also be examined using this approach. If the removal part of an enhancer repressed its activity, it would suggest the entire sequence of the enhancer was important in its function and therefore followed the enhanceosome model with strict and additive arrangement of TFBS and cognate TF.

The putative TAD outlined in Chapter 4.4 and chromatin topology within this region could also be investigated using CRISPR-Cas9. The CTCF and cohesin sites at the predicted TAD boundary could be deleted, as has previously been demonstrated to perturb enhancer function (Guo *et al.* 2015). This could cause fidelity in *Ccn2* enhancer-*Ccn2* promoter interaction to be lost, and therefore the temporospatial specificity in *Ccn2* transcription to be lost, resulting in aberrant patterns of gene activation or repression; which could have pathological

consequences (Lupiáñez *et al.* 2015). Moreover, elucidation of the physiological role of *Ccn2* enhancers through genomic manipulation could in turn, aid the understanding and amelioration of the profound diseases caused by aberrant *Ccn2* expression such as osteoarthritis.

4.5 Conclusions

Enhancer function has been proven to critically regulate the transcription of several skeletogenesis related genes, especially during embryonic development. For example, Chen *et al*. (2016) found multiple enhancers of *GDF5* expression that discretely specify gene transcription localisation, with elements driving expression in separate skeletal tissues. With the discovery of one of the largest distance enhancer-promoter interactions, Lettice *et al*. (2003) found that an enhancer located 1mb away from target gene *SHH* functions in limb patterning, and that mutations within this sequence are linked with polydactyly phenotype. This highlights the importance of the current findings with *Ccn2* enhancers mediating differential transcription of *Ccn2* in skeletogenesis, especially given the critical role of CCN2 in physiological chondrocyte behaviour during development and beyond.

Chondrocyte based expression of *Ccn2* can be driven by the -137kb, -148kb and -198kb enhancers at E15.5. Within articular chondrocytes, where transgene expression was observed to be driven by all three enhancers, enhancer DNA-cell lineage-specific TF interaction, further activator protein-protein interaction and chromatin looping through CTCF and cohesin, could culminate in the enhancers being active and brought into close proximity with the promoter region for *Ccn2*. This would lead to increased transcriptional output from *Ccn2* in articular chondrocytes specifically, as illustrated in Figure 4.13. The utilisation of enhancers could also occur sequentially with enhancer poising, in order to ensure robust expression of *Ccn2* throughout skeletogenesis, in both chondrocyte and subsequently osteoblastic cells.

The endogenous expression of Ccn2 therefore reflects composite activities of several enhancers, which are capable of functioning concurrently yet in highly specific cell and tissue types in order to confer temporospatial specificity in the transcription of Ccn2.

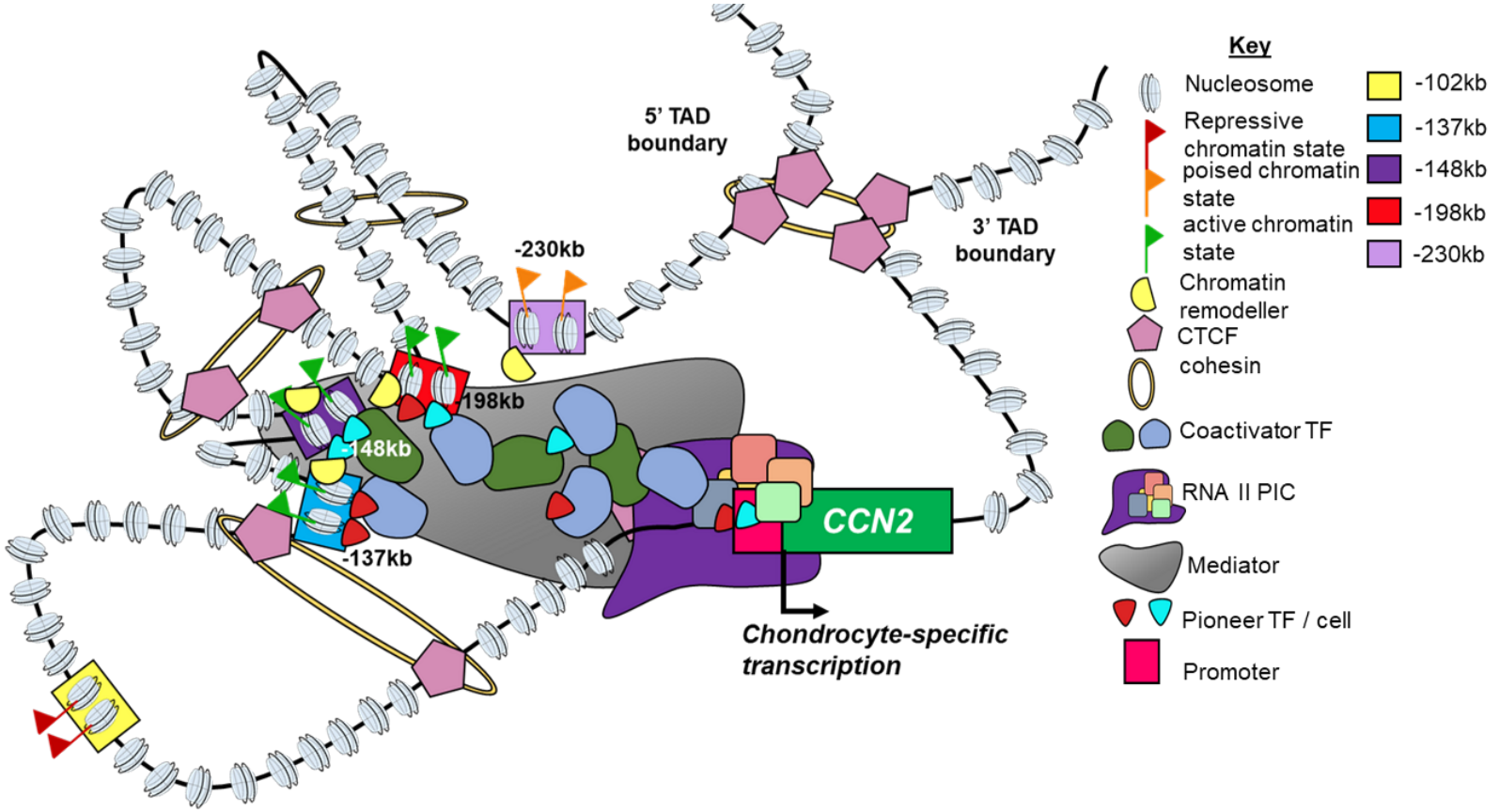


Figure 4.13: Schematic representation of the TAD of approximately 300kb that *Ccn2* may reside in, and the TF-enhancer and protein interactions that determine -137kb, -148kb and -198kb activity within chondrocyte cells. Enhancers are active in a highly temporospatial manner, reflecting specific, open chromatin structure that enables interaction with pioneer and cell lineage-specific transcription factors. Further protein-protein interactions between TF and transcriptional coactivators and mediator protein complex, in addition to interaction between CTCF and cohesin with target sites facilitates chromatin looping that brings about interaction between the -137kb, -148kb, -198kb enhancers and *Ccn2* promoter region in the nucleus of articular chondrocytes. This culminates in the assembly of the pre-initiation complex and protein machinery required for transcription at the TSS, therefore enhancing expression of *Ccn2.*

5. Appendices

5.1 Datasets assimilated into the UCSC Genome Browser

Datasets available as part of the UCSC Genome Browser (Kent *et al.* 2002) were used in the first instance in order to identify putative enhancers of *Ccn2* predominantly within limb tissues.

characteristic	source	First author/ Credit	Track id	Sample	Time-point
H3K4me3	ENCODE	ENCODE/LICR	Limb H3K4m3	WT C57BL/6 Limb	E14.5
H3K4me1	ENCODE	ENCODE/LICR	Limb H3K4m1	WT C57BL/6 Limb	E14.5
H3K27ac	ENCODE	ENCODE/LICR	Limb H3K27a	WT C57BL/6 Limb	E14.5
DNase I	ENCODE	ENCODE /UW	Fore limb bud S 1	WT CD-1 fore limb bud	E11.5
DNase I	ENCODE	ENCODE /UW	Hind limb bud S 1	WT CD-1 fore hind bud	E11.5
DNase I	ENCODE	ENCODE /UW	Fibroblast S 1	WT C57BL/6 lung	8 weeks
DNase I	ENCODE	ENCODE /UW	Mesoderm S 1	WT CD-1 Mesoderm	E11.5
DNase I	ENCODE	ENCODE /UW	NIH-3T3 S 1	NIH3T3 immortalised fibroblast cells	n/a

Table 5.1: Datasets used concerning chromatin characteristics associated with enhancers used to predict enhancers of Ccn2 within limb tissue

The following GEO-DataSets were used to examine TF-DNA interaction and potential proteins that regulated enhancer function (Table 5.2).

Protein	credit/ citation	Series Accession	Sample accession	Sample	Time-point
SOX9	Garside *et al.* (2015)	GSE73225	GSM1888973	WT limb	E12.5
	Ohba *et al.* (2015)	GSE69109	GSM1692996	WT, rib chondrocyte	P1
			GSM1693007	WT nasal chondrocyte	E17.5
Runx2	Meyer *et al.* (2014)	GSE41920	GSM1027478	MC3T3-E1, undifferentiated	n/a, cell line
			GSM1027496	MC3T3-E1, osteogenic differentiation	n/a, cell line
Jun	He *et al.*(2016)	GSE73372	GSM1891979	WT, rib chondrocyte	P1
p300	Chronis *et al.* (2017)	GSE90893	GSM2417166	Mouse embryonic fibroblast	n/a
CTCF	Barutcu *et al.*(2014)	GSE55045	GSM1328456	MC3T3-E1, day 0 osteogenic differentiation	n/a, cell line
			GSM1328457	MC3T3-E1, day 9 osteogenic differentiation	n/a, cell line
			GSM1328458	MC3T3-E1, day 28 osteogenic differentiation	n/a, cell line
H3K27ac	Infante *et al.* (2015)	GSE64055	GSM1563530	WT ICR strain, forelimb, replicate 1	E11.5
			GSM1563531	WT ICR strain, forelimb, replicate 2	
			GSM1563533	WT ICR strain, hind limb, replicate 1	
			GSM1563534	WT ICR strain, hind limb, replicate 2	
cohesin	(Kim *et al.* 2018)	GSE81676	GSM2169995	Mouse embryonic stem cell line ESC-46c	n/a

Table 5.2: ChIP-Seq datasets used from the GEO-dataset resource

5.2 Plasmid maps

5.2.1 Hsp68LacZGW vector maps

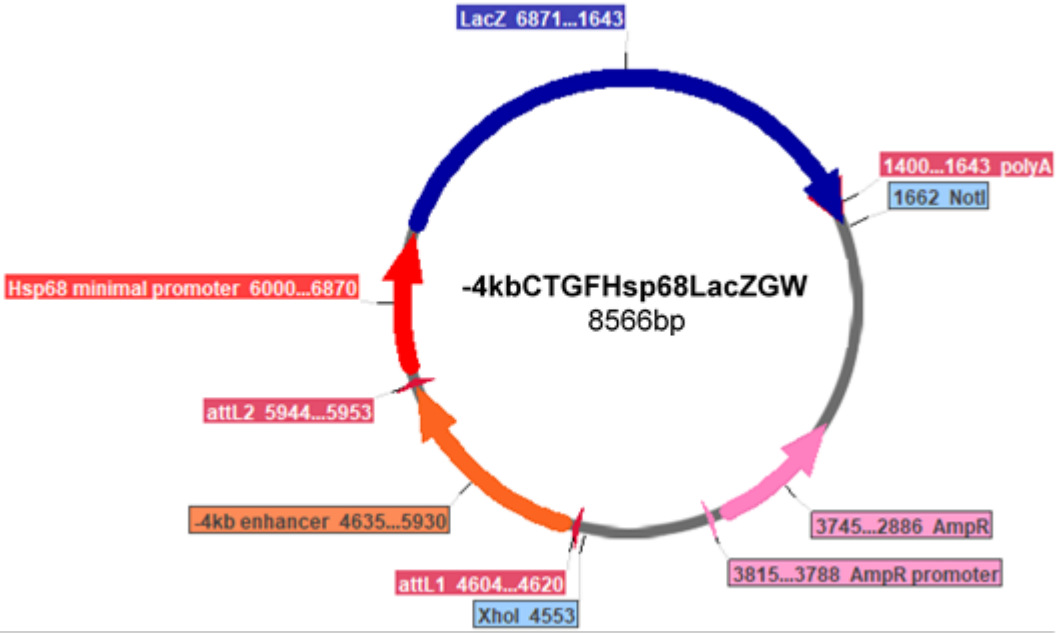


Figure 5.1: -4kbCTGFHsp68LacZGW plasmid map. The 1295bp -4kb enhancer (orange) drives the expression of LacZ reporter gene (blue) via an Hsp68 minimal promoter (red). The plasmid also contains sequences necessary for propagation and selection in E.coli. This plasmid was linearised using XhoI and NotI restriction enzymes.

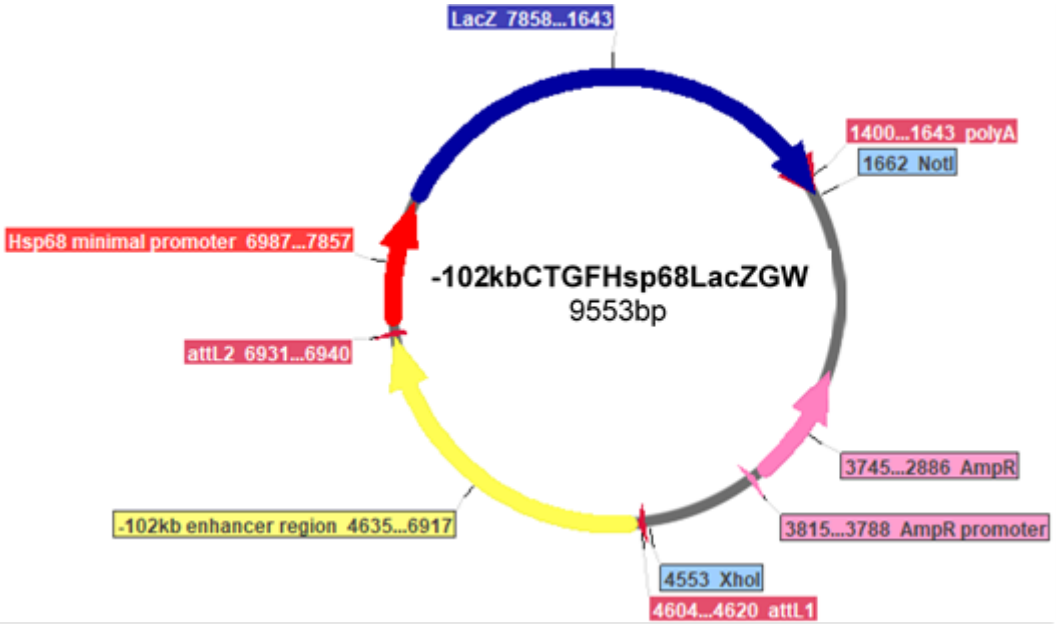


Figure 5.2: -102kbCTGFHsp68LacZGW plasmid map. The -102kb enhancer (2282bp) (yellow) regulates the transcription of *LacZ* reporter gene (blue) via an Hsp68 minimal promoter (red). The plasmid also contains sequences necessary for propagation and selection in *E.coli*. This plasmid was linearised using XhoI and NotI restriction sites

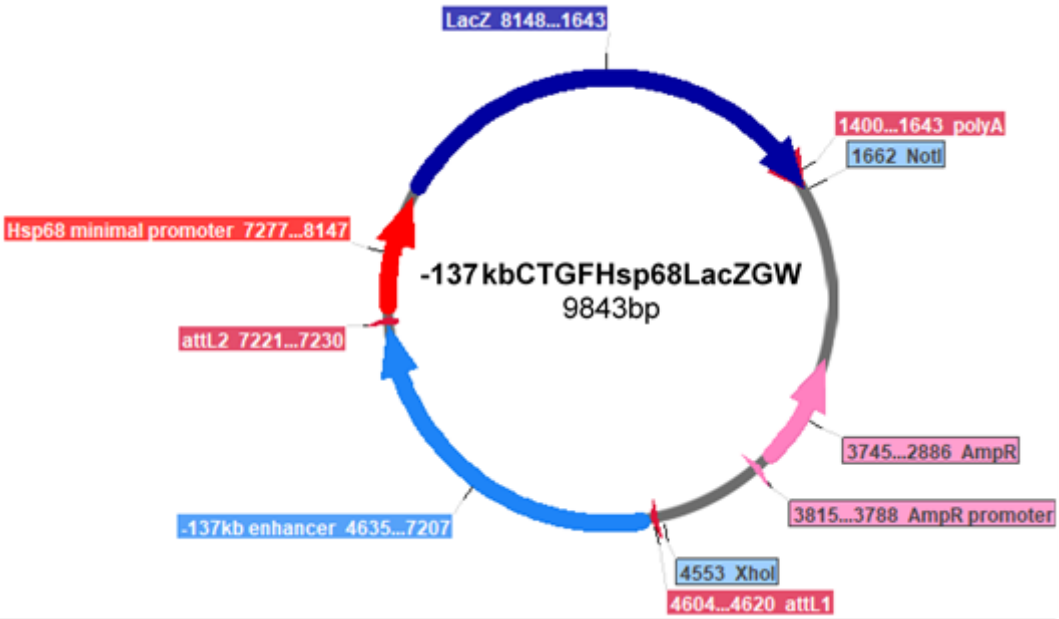


Figure 5.3: -137kbCTGFHsp68LacZGW plasmid map. The full -137kb enhancer region (2572bp) (pale blue) drives the expression of *LacZ* (dark blue) via an Hsp68 promoter (red). The plasmid also contains sequences necessary for propagation and selection in *E.coli*. This plasmid was linearised using XhoI and NotI restriction enzyme sites

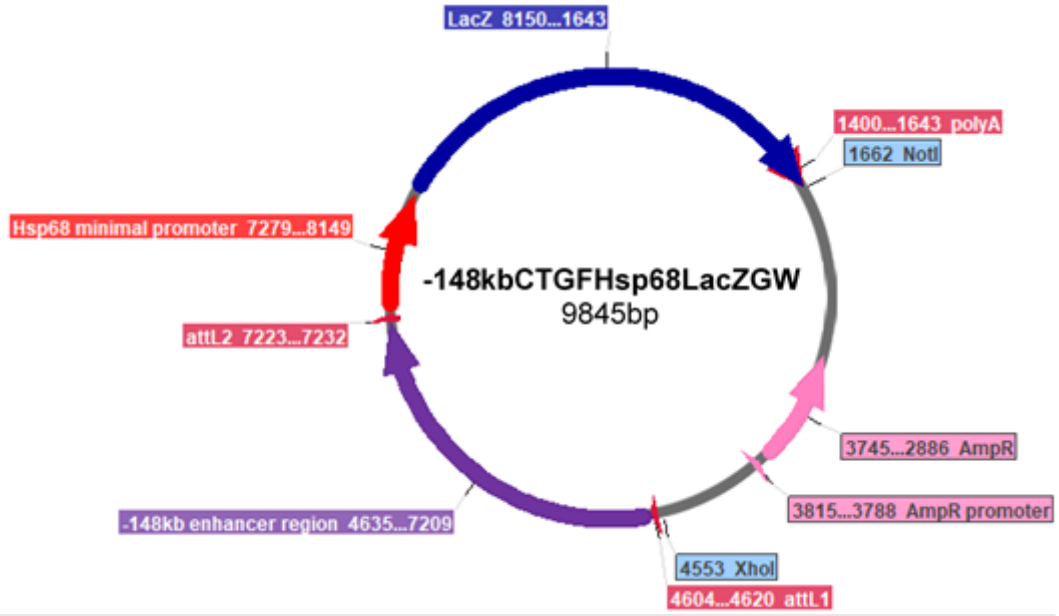


Figure 5.4: -148kbCTGFHsp68LacZGW plasmid map. The full -148kb enhancer region (purple) drives the expression of a *LacZ* reporter gene (blue) via an Hsp68 promoter. The plasmid also contains sequences necessary for propagation and selection in *E.coli.* XhoI and NotI restriction enzymes were used to linearise this plasmid.

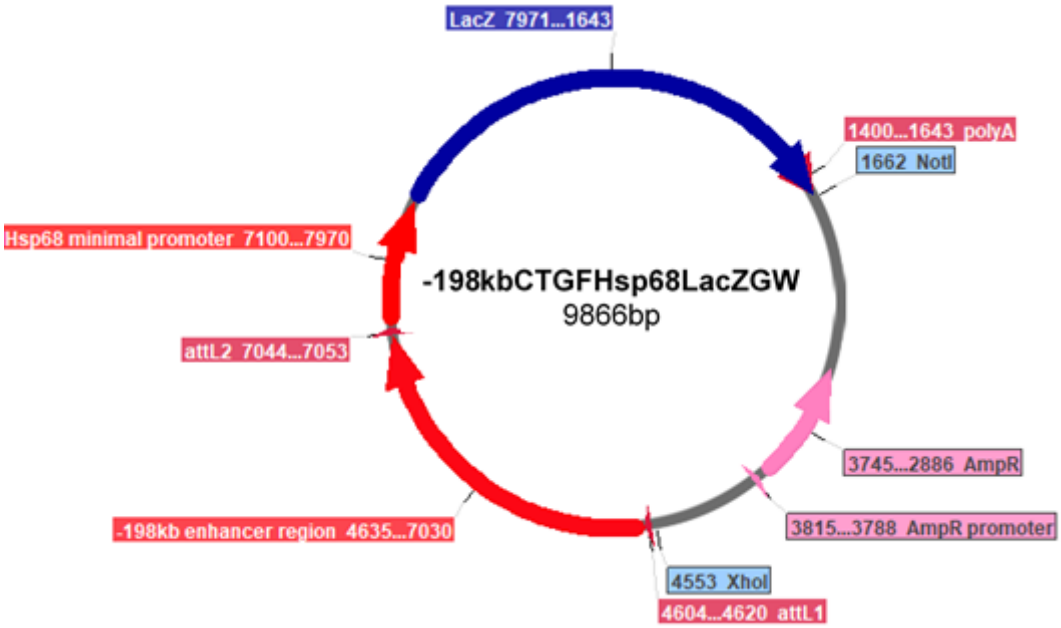


Figure 5.5:-198kbCTGFHsp68LacZGW plasmid map. The -198kb enhancer sequence (2395bp) controls *LacZ* reporter gene expression via an Hsp68 minimal promoter region. The plasmid also contains sequences necessary for propagation and selection in *E.coli.* Plasmid linearisation was carried out using XhoI and NotI.

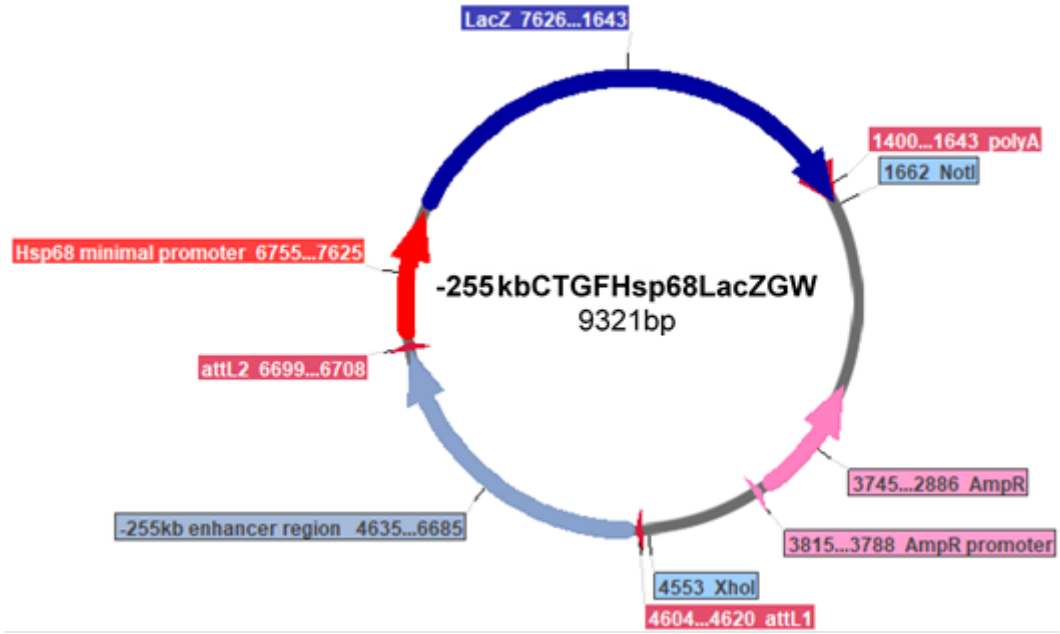


Figure 5.6: -255kbCTGFHsp68LacZGW plasmid map. The -255kb enhancer region (2050bp) (pale blue) regulates the expression of *LacZ* reporter gene (dark blue) via an Hsp68 minimal promoter (red). The plasmid also contains sequences necessary for propagation and selection in *E.coli.* This plasmid was linearised using XhoI and NotI

5.2.2 Maps of *LacZ* plasmids without Gateway®

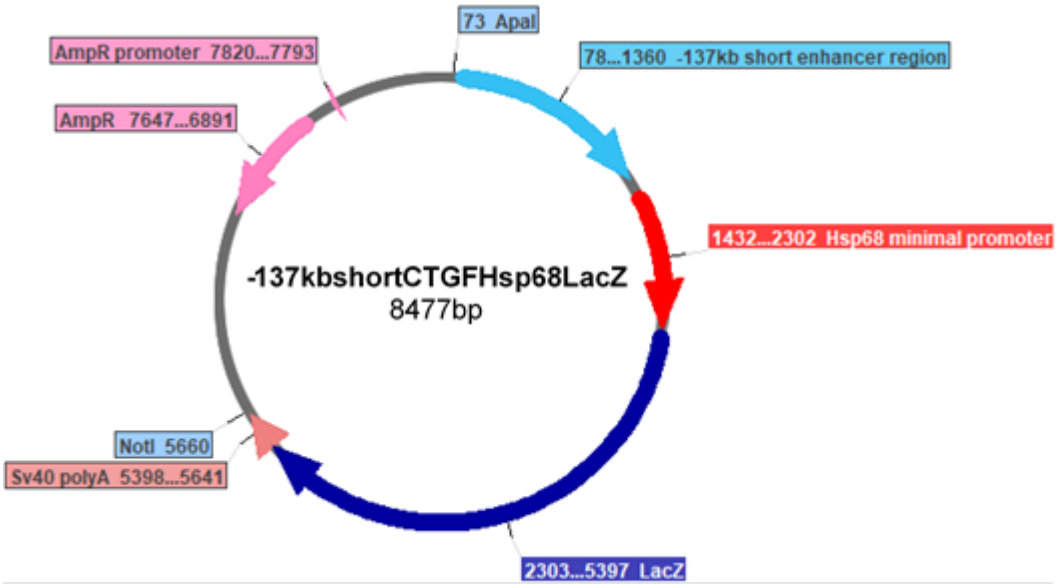


Figure 5.7: -137kbshortCTGFHsp68LacZ plasmid map. The shorter form of the -137kb enhancer, with sequence spanning the most conserved region (1275bp) (bright blue) controls the expression of *LacZ* reporter gene (dark blue) via an Hsp68 minimal promoter region (red). The plasmid also contains sequences necessary for survival and propagation in *E.coli.* ApaI and NotI restriction enzymes were used to linearise this plasmid.

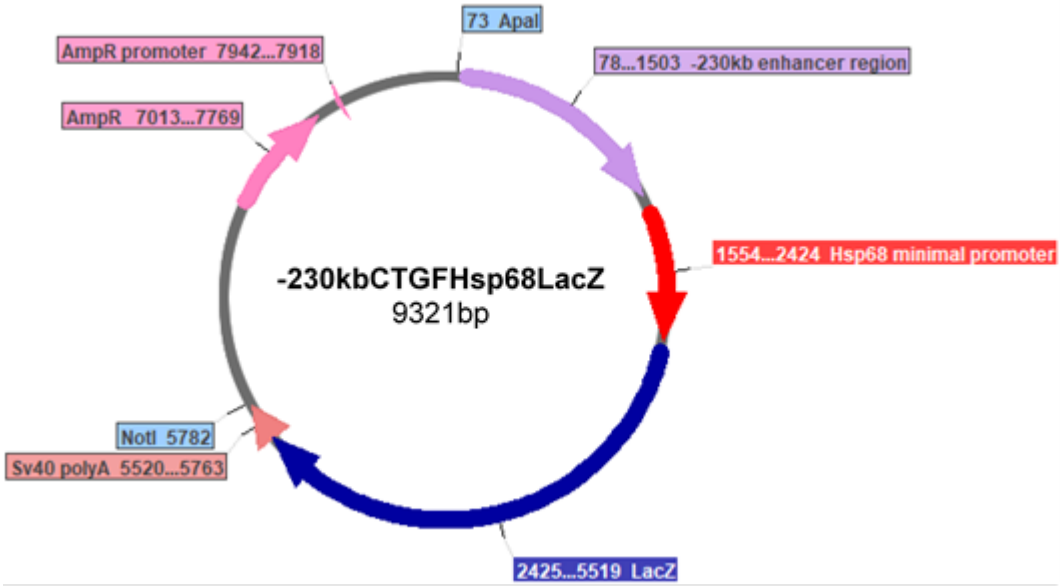


Figure 5.8: -230kbCTGFHsp68LacZ plasmid map.

The -230kb enhancer region (1423bp) (lilac) drives the expression of *LacZ* reporter gene (blue) via an Hsp68 minimal promoter (red). The plasmid also contains sequences required for propagation and survival in *E.coli*. This plasmid was linearised using ApaI and NotI restriction enzyme sites.

5.2.3 Maps of pSLF01 enhancer containing variant plasmids

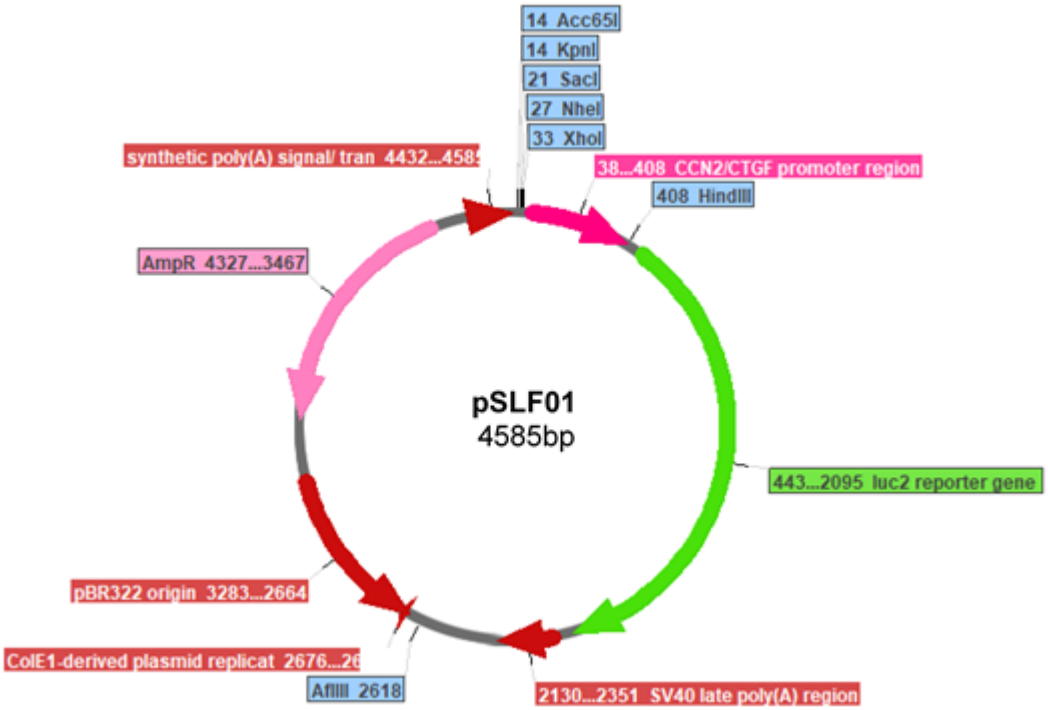


Figure 5.9: pSLF01 plasmid map. The promoter region spanning from -351bp to +17bp relative to *CCN2* was cloned into the pGL4.10 vector. The *CCN2* promoter region regulates the expression of firefly *Luc2* luciferase reporter gene and was cloned into the MCS using XhoI and HindIII. The plasmid also contains sequences necessary for propagation in E.coli such as origin of replication and Ampicillin resistance gene for selection.

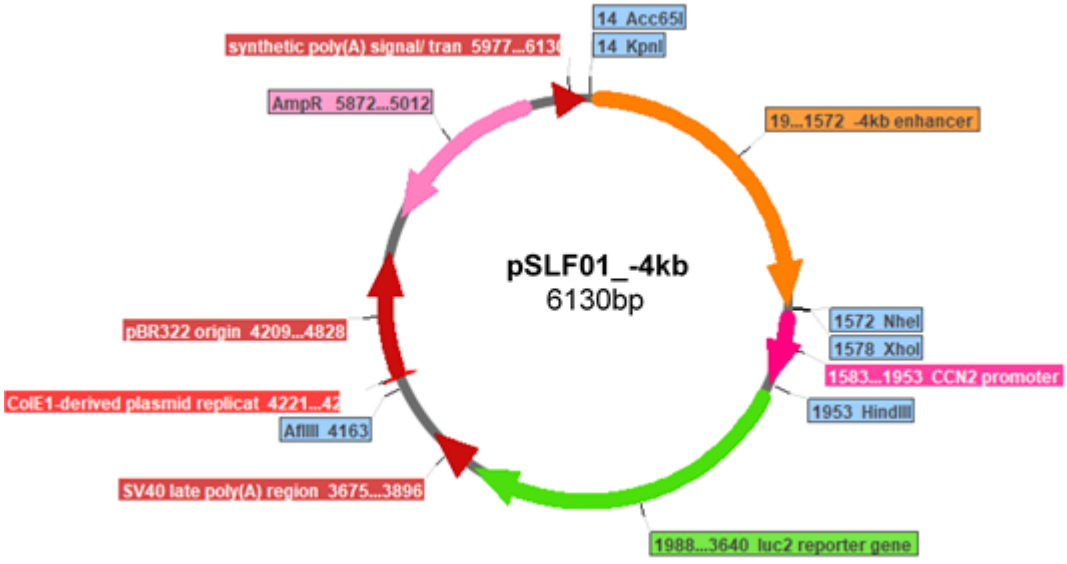


Figure 5.10: pSLF01_-4kb plasmid map. The -4kb enhancer region (orange) was cloned into pSLF01 using KpnI and NheI sites and regulates the expression of *Luc2* firefly luciferase reporter gene (green) in conjunction with the *CCN2* promoter region (bright pink). This plasmid also contains sequences necessary for survival and selection within E.coli. This plasmid was linearised using the Acc65I and AflIII sites

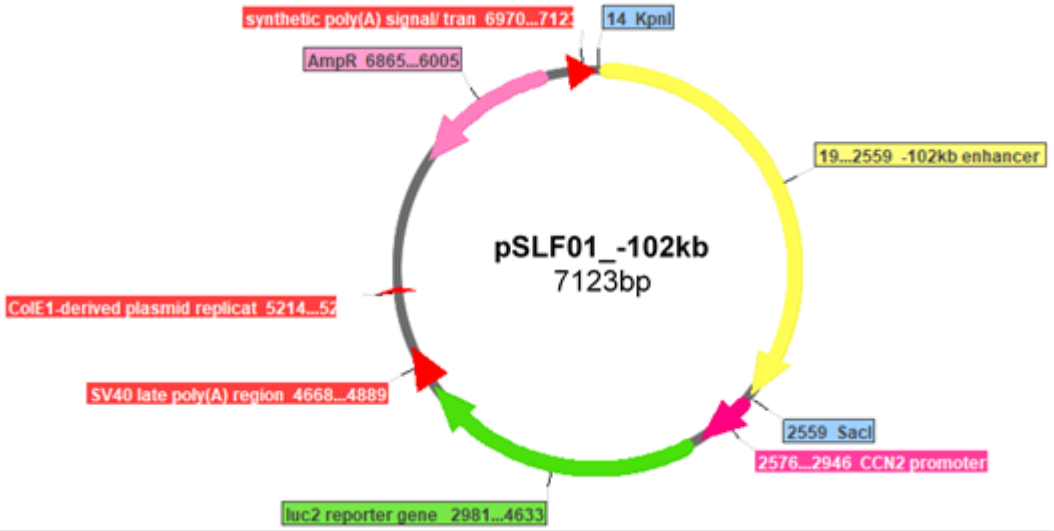


Figure 5.11: pSLF01_-102kb plasmid map. The -102kb enhancer (yellow) was cloned into pSLF01 using KpnI and SacI sites. The enhancer regulates the expression of firefly luciferase luc2 expression (green) via the *CCN2* promoter (pink).

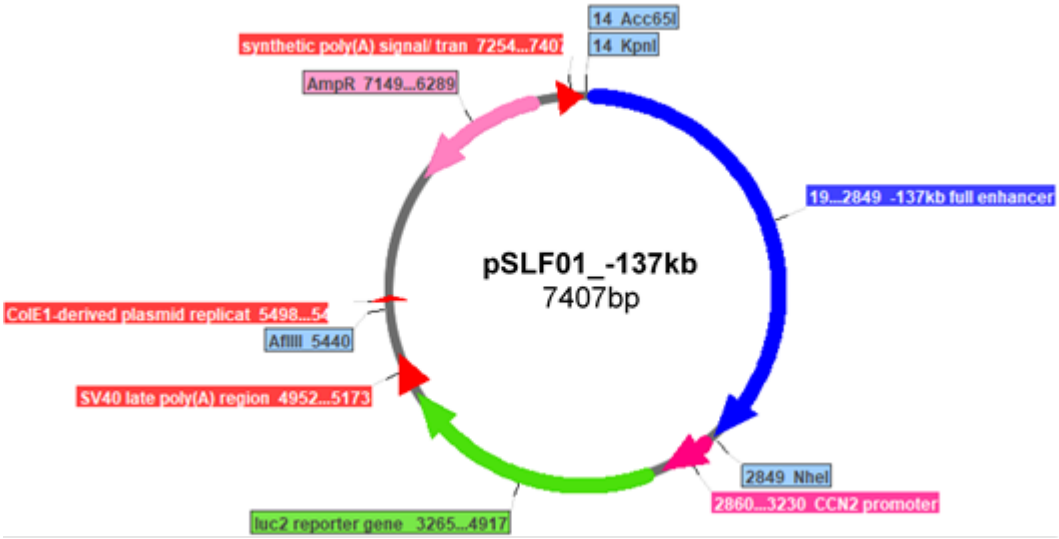


Figure 5.12: pSLF01_-137kb plasmid map. The full -137kb enhancer region (blue) was cloned into pSLF01 using KpnI and NheI. The enhancer controls expression of *luc2* firefly luciferase reporter gene (green) via the *CCN2* promoter region (pink). The plasmid also contains necessary sequences for propagation and selection in *E.coli*.

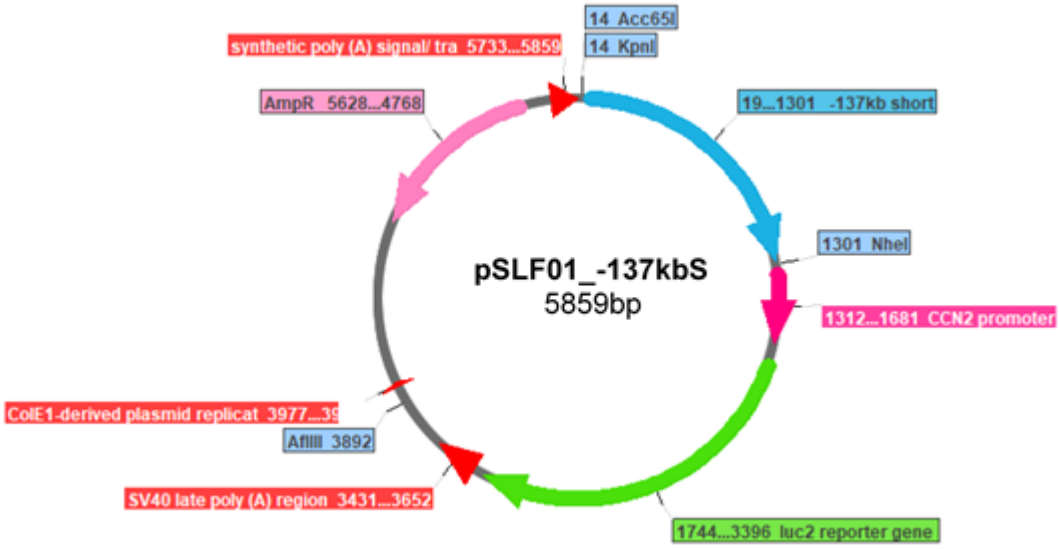


Figure 5.13: pSLF01_-137kb_S plasmid map. The short version of the -137kb enhancer, composed of the most evolutionarily conserved sequence regulates firefly luciferase *luc2* expression via the *CCN2* promoter region. The plasmid also contains sequences necessary for propagation and selection in *E.coli*

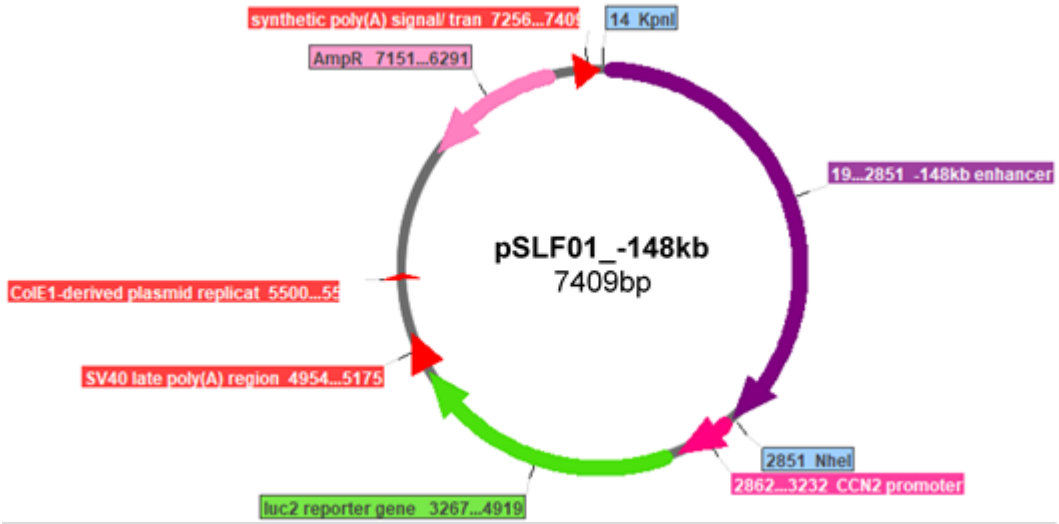


Figure 5.14: pSLF01_-148kb plasmid map. The -148kb enhancer (purple) drives the expression of firefly luciferase reporter gene (green) via the *CCN2* promoter (pink). The enhancer was cloned into pSLF01 using KpnI and NheI sites. The plasmid also contains sequences necessary for propagation and selection in *E.coli*.

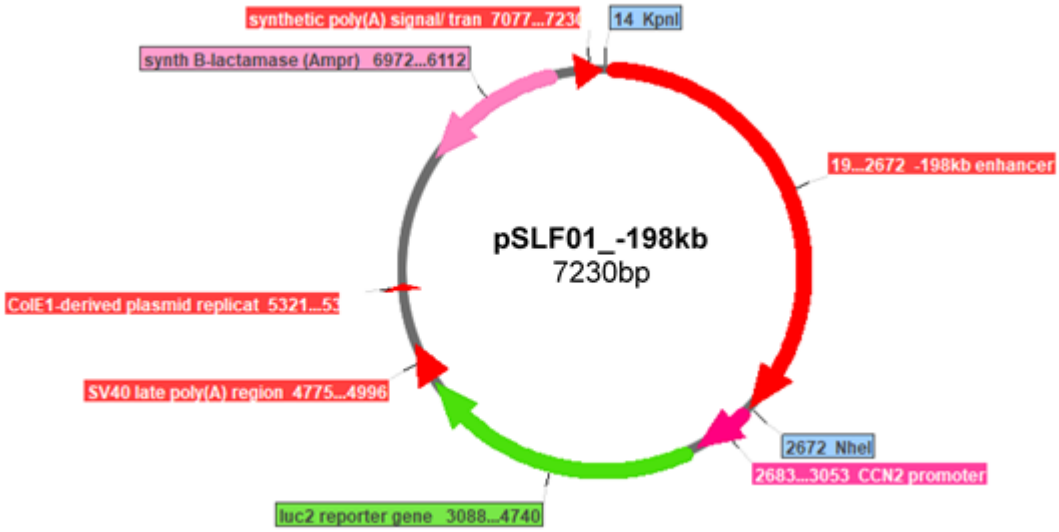


Figure 5.15: pSLF01_-198kb plasmid map. The -198kb enhancer (red) was cloned into pSLF01 using KpnI and NheI sites controls the expression of *luc2* luciferase reporter gene (green) via the CCN2 promoter region (pink). The plasmid also contains sequences for propagation and selection in *E.coli*.

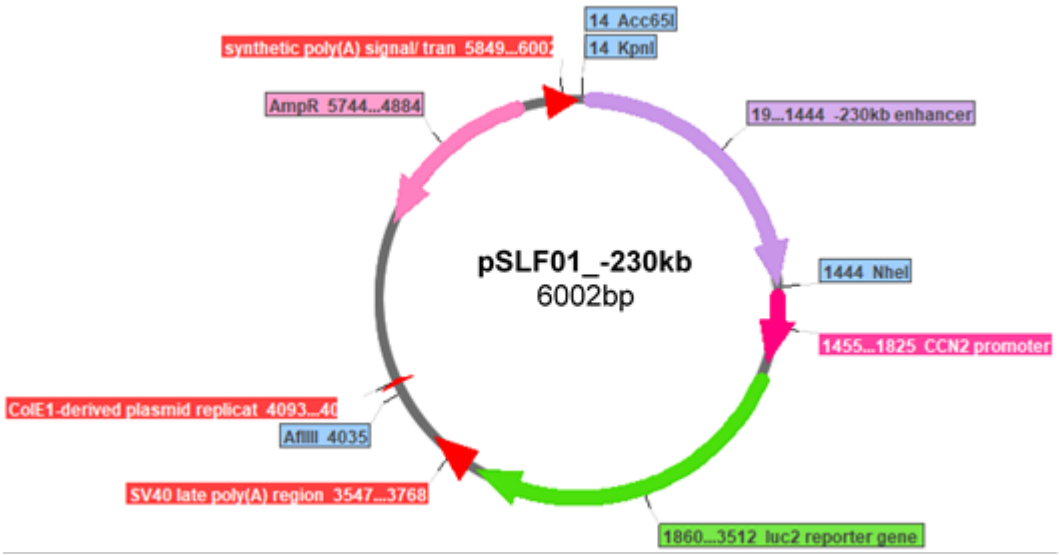


Figure 5.16: pSLF01_-230kb plasmid map. The -230kb enhancer (lilac) was cloned into pSLF01 using KpnI and NheI sites. The enhancer drives the expression of firefly luciferase reporter gene (green) via the *CCN2* promoter. The plasmid also contains sequences for propagation and selection in *E.coli.*

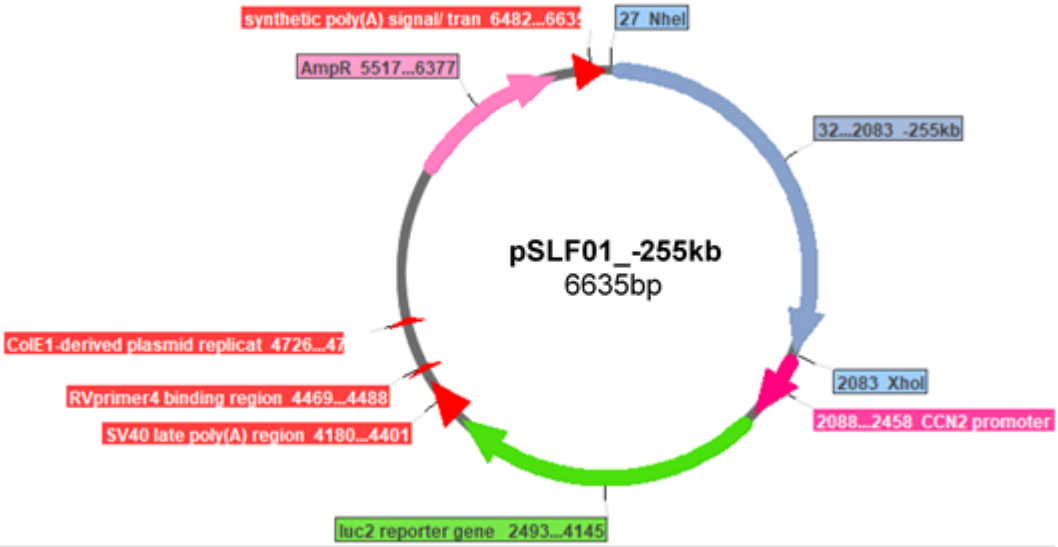


Figure 5.17: pSLF01_-255kb plasmid map. The -255kb enhancer region was cloned into pSLF01 using NheI and XhoI sites. This enhancer drives the expression of firefly luciferase (*luc2*, green) via the CCN2 promoter. This plasmid also requires sequences necessary for propagation and selection of plasmid in *E.coli.*

5.3 CRISPR

The following Tables and Figure supplement Chapters 2.4 and 3.9 regarding the CRISPR-Cas9 based manipulation of *CCN2* enhancer sequences.

Enhancer	Boundary	Target range	Guide	Guide sequence	Off-target locus
-137kb	5’	chr10:24457493-24457893	137_5’ _1	CTGGTATCCACAATGAGGTTCGG	chr19:53809042
			137_5’ _2	TGGTATCCACAATGAGGTTCGGG	chr19:53809043 and chr2:15783355
	3’	chr10:24460063-24460463	137_3’ _1	TATGGAGTTTTGCTGAACTGTGG	chr8:14881264
			137_3’ _2	GTTGACTCTAAACTATTATGAGG	n/a
-148kb	5’	chr10:24446683-24447083	148_5’ _1	ATGTCCCTAACCCTTGCTTATGG	n/a
			148_5’ _2	TATGCCATAAGCAAGGGTTAGGG	chr14:37696053
	3’	chr10:24448955-24449955	148_3’ _1	CTCAGGACTACAATTGCATCAGG	chr11:53796130
			148_3’ _2	CCGTCTTACAAAACATATAATGG	chr15:43848059
-148kb/-137kb	5’	chr10:24446000-24461000	148/137_5’	TTGATAGTGGCCACAAGAGAGGG	n/a
	3’	chr10:24449000-24463000	148/137_3’	GTGAATCCCAAATTGTGACTCGG	n/a
-230kb	5’	chr10:24364451-24364851	230_5’ _1	AGTTACCATAAGCCCACCAGAGG	n/a
			230_5’ _2	TAAGGCTTTACATTAGCGAATGG	n/a
	3’	chr10:24365949-24366349	230_3’ _1	AGCACAGATGATGTCGTTTGCGG	n/a
			230_3’ _2	AAGAGCAATGATGTGCTATACGG	n/a

Table *5.3:* Guide RNA identified for each enhancer of interest and genomic range which was used to find them

Guide	Forward primer	Reverse primer	Product size (bp)	Approximate fragment sizes (bp)
-137kb_5' _1 alone	137_5'C_F	137_5'C_R	529	200 and 300
	-137CTGF Fw	-137CTGF Rv	2572	180 and 2400
-137kb_5' _2 alone	137_5'C_F	137_5'C_R	529	200 and 300
	-137CTGF Fw	-137CTGF Rv	2572	180 and 2400
-137kb_3' _1 alone	137_3' /2_C_F	137_3' /2_C_R	978	425 and 550
-137kb_3' _2 alone	-137CTGF Fw	-137CTGF Rv	2572	130 and 2440
-137kb_5' 1 and -137kb_3' 1	-137CTGF Fw	-137CTGF Rv	2572	130,180 and 2260
	-137CTGF Fw	137_3' /2_C_R	2864	180,430,2260
-137kb_5' 1 and -137kb_3' 2	-137CTGF Fw	137_3' /2_C_R	2864	180,250 and 2450
-137kb_5' 2 and -137kb_3' 1	-137CTGF Fw	-137CTGF Rv	2572	130, 200 and 2260
	-137CTGF Fw	137_3' /2_C_R	2864	200, 430 and 2260
-137kb_5' 2 and -137kb_3' 2	-137CTGF Fw	137_3' /2_C_R	2864	180, 250 and 2450

Table *5.4:* primer combinations used for *in vitro* validation of -137kb targeting gRNA function and fragment resultant from successful cleavage of PCR substrate

Guide	Forward primer	Reverse primer	Product size (bp)	Approximate fragment sizes (bp)
-148kb_5' _1 alone	-148CTGF FW	-148CTGF Rv	2574	160 and 2420
	-148CTGF Fw	148_3'/1_C_R	3012	160 and 2850
-148kb_5' _2 alone	-148CTGF FW	-148CTGF Rv	2574	150 and 2430
	-148CTGF FW	148_3'/1_C_R	3012	150 and 2860
-148kb_3' _1 alone	148_3'/1_C_F	148_3'/1_C_R	752	300 and 450
-148kb_3' _2 alone	-148CTGF FW	-148CTGF Rv	2574	450 and 2130
-148kb_5' 1 and -148kb_3' 1	-148CTGF Fw	148_3'/1_C_R	3012	160, 300 and 2540
-148kb_5' 1 and -148kb_3' 2	-148CTGF FW	-148CTGF Rv	2574	160, 450 and 1970
	-148CTGF Fw	148_3'/1_C_R	3012	160, 890 and 1970
-148kb_5' 2 and -148kb_3' 1	-148CTGF Fw	148_3'/1_C_R	3012	150, 300 and 2560
-148kb_5' 2 and -148kb_3' 2	-148CTGF FW	-148CTGF Rv	2574	150,450 and 1980
	-148CTGF Fw	148_3'/1_C_R	3012	150,900 and 1980

Table *5.5:* primer combinations used for *in vitro* validation of –148kb targeting gRNA function and fragment resultant from successful cleavage of PCR substrate

Guide	Forward primer	Reverse primer	Product size (bp)	Approximate fragment sizes (bp)
-230kb_5' _1 alone	-230kb F	-230kb R	1500	190 and 1310
	-230kb F	230_3'_C_R	1950	200 and 1750
-230kb_5' _2 alone	230_5'_C_F	230_5'_C_R	869	250 and 620
-230kb_3' _1 alone	230_3'_C_F	230_3'_C_R	648	310 and 340
-230kb_3' _2 alone	230_3'_C_F	230_3'_C_R	648	240 and 410
-230kb_5' _1 and -230kb_3' _1	-230kb F	230_3'_C_R	1950	190, 320 and 1450
-230kb_5' _1 and -230kb_3' _2	-230kb F	230_3'_C_R	1950	190, 240 and 1520
-230kb_5' 2 and -230kb_3' _1	230_5'_C_F	230_3'_C_R	2630	320, 600 and 1700
230kb_5' 2 and -230kb_3' _2	230_5'_C_F	230_3'_C_R	2630	240, 600 and 1780

Table *5.6* primer combinations used for *in vitro* validation of –230kb targeting gRNA function and fragment resultant from successful cleavage of PCR substrate

CRISPR *in vitro* digestion were optimised with comparison between Cas9 manufacturer Toolgen's recommended assay conditions and those recommended in previously published papers (Gong *et al.* 2018a; Kouranova *et al.* 2016) as represented in Figure 5.18.

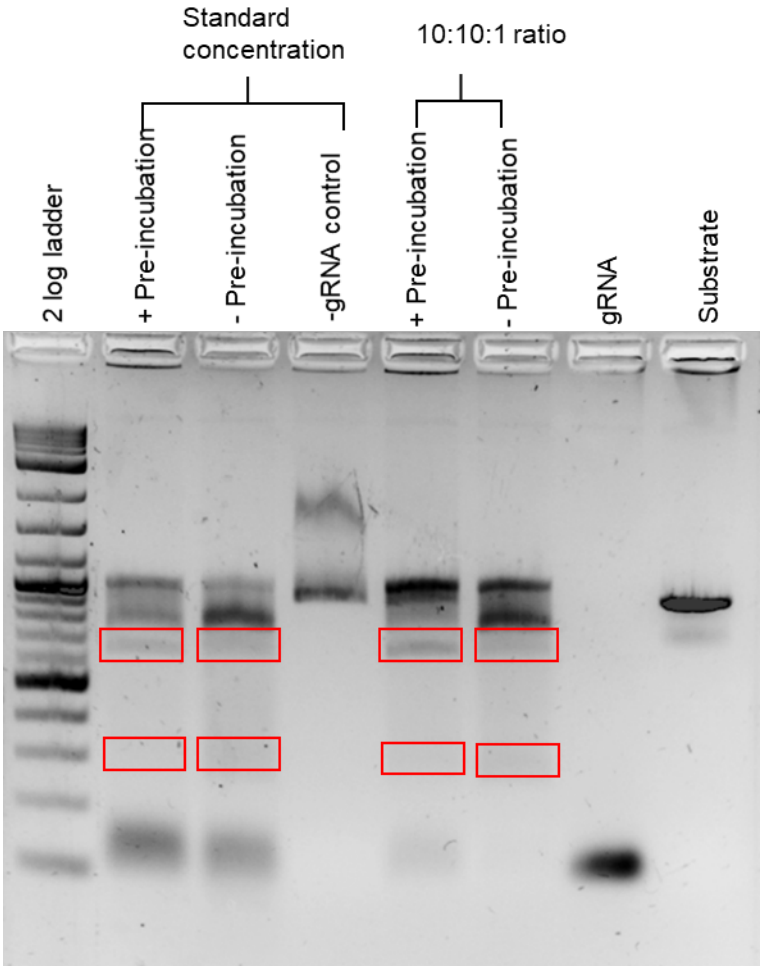


Figure 5.18: Optimisation of CRISPR gRNA-Cas9 complexing and cutting conditions. The -230kb 5'_2 guide RNA was used to cut -230kb 5'_C_Fx -230kb 3'_R PCR substrate. The use of 10:10:1 Cas9 protein: gRNA: substrate molar ratio led greater yield of fragments corresponding to the cutting of the fragment mediated by the gRNA and Cas9 of 300bp and 800bp.

5.4 Examination of transcription factor binding sites

5.4.1 TRAP prediction of TFBS

The following tables and figures concern the examination of transcription factor binding sites within the enhancers of *CCN2* transcription and oligonucleotides used for EMSA. Transcription factors linked to the lineages in which the enhancer functioned *in vivo* are highlighted in each Table concerning prediction of TFBS.

Rank	P value	Matrix ID	TF
1	0.00302756026148676	M00191	ER
2	0.00306086699554986	M00312	BEL1
3	0.00813297102955457	M0050	E2F
4	0.0087941028390065	M01232	SATB1
5	0.00974567996503806	M00738	E2F4DP1
6	0.0107435121100254	M01301	MEF2
7	0.0130352706297904	M00102	CDP
8	0.0132417173548243	M00740	E2F1DP1RB
9	0.0133360645387725	M00423	FOXJ2
10	0.0139115501333639	M01148	DMRT3
11	0.014544328859379	M00340	ETS2
12	0.0147656854963978	M00425	E2F
13	0.0177211119941748	M00426	E2F
14	0.0202980354370221	M00427	E2F
15	0.021083873832637	M00088	IK3
16	0.0215978097726057	M00941	MEF2
17	0.0216922567354194	M00769	AML
18	0.0221050523271429	M01150	DMRT5
19	0.0228993941888092	M00406	MEF2
20	0.0233318962508079	M00212	POLY_C
21	0.0235126080211755	M00211	PADS_C
22	0.0250718825354077	M00347	GATA1
23	0.025211476876477	M00103	CLOX
24	0.0257935119622424	M01163	ELK1
25	0.025968099177357	M00078	EVI1
26	0.0275298736081423	M00231	MEF2
27	0.0276094436361739	M00278	LMO2COM
28	0.0279624020302909	M00232	MEF2
29	0.0286464424476612	M00216	TATA
30	0.0305451939520733	M00420	MEISAHOXA9
31	0.0309182548643682	M00410	SOX9
32	0.0327715656653854	M00953	AR
33	0.0328356796113028	M01145	CMYC
34	0.0330845100556741	M01149	DMRT4
35	0.0368864978222559	M00955	GR
36	0.0381678754686549	M00085	ZID
37	0.0384161235299434	M00493	STAT5A
38	0.0386322024944555	M00006	MEF2
39	0.0427793122565435	M01165	ELK1
40	0.0428422793267728	M00214	ESF1
41	0.0428592794653477	M00939	E2F1
42	0.0431017031466299	M00918	E2F1
43	0.0434606048155622	M00118	MYCMAX
44	0.0470501352211926	M00481	AR
45	0.0472826370499406	M00119	MAX
46	0.0478701697973889	M00799	MYC
47	0.0483058286405239	M00954	PR

48	0.0498649571840802	M01249	HIF2A

Table 5.7: TRAP predicted TFBS within the -102kb enhancer region. The whole -102kb enhancer sequence was inputted into TRAP. The motifs with P≤0.05, therefore significant prediction of binding interaction are listed.

Rank	P value	Matrix ID	TF
1	8.7785963709397e-07	M00750	HMGIY
2	2.64434851466877e-05	M00493	STAT5A
3	0.0020805236051108	M00515	PPARG
4	0.00495926452064788	M01150	DMRT5
5	0.00534351999516525	M01185	BCL6
6	0.00617540437290076	M00026	RSRFC4
7	0.00651611021625731	M00407	RSRFC4
8	0.00675490046708382	M00231	MEF2
9	0.0070591965752157	M01023	HSF1
10	0.012547037893872	M00339	ETS1
11	0.0133322201587713	M01224	P50RELAP65
12	0.0140390771575355	M00215	SRF
13	0.0157185149860948	M00771	ETS
14	0.0162037091035698	M01112	RBPJK
15	0.0172400333364215	M01125	OCT4
16	0.0201924172322243	M00619	ALX4
17	0.0233462958115297	M01123	NANOG
18	0.0234274651574943	M00926	AP1
19	0.0242734937246825	M00960	PR
20	0.025129216568412	M01267	FRA1
21	0.0251998539002951	M01244	HSF2
22	0.0260325898804088	M01486	DLX7
23	0.0282090302099872	M00212	POLY_C
24	0.0307723028560771	M00097	PAX6
25	0.0313132122814407	M01403	OTX3
26	0.031335976692661	M00925	AP1
27	0.0333899378425998	M00291	FREAC3
28	0.0334661983420653	M00410	SOX9
29	0.0387491217299648	M00025	ELK1
30	0.0389501544033724	M01398	SIX6
31	0.0399991472699811	M00340	ETS2
32	0.0404773485433498	M00203	GATA
33	0.0405726937449582	M00194	NRKB
34	0.0412225909590964	M00057	COMP1
35	0.0412861307458219	M00655	PEA3
36	0.0416275341462234	M00955	GR
37	0.0430102373517307	M01447	PITX2
38	0.0431915302460939	M00232	MEF2
39	0.049054271502528	M00512	PPARG

Table 5.8: TRAP predicted TFBS within the -137kb full enhancer region. The motifs with P≤0.05, therefore significant prediction of binding interaction are listed. Chondrocyte

related AP-1 and SOX9 sites are highlighted (He *et al.* 2016). The SOX9 site was validated in Chapter 3.8.

Rank	P value	Matrix ID	TF
1	0.000358596885802642	M00410	SOX9
2	0.000855009838309551	M01244	HSF2
3	0.000957658735226574	M01023	HSF1
4	0.00185888133831547	M00926	AP1
5	0.0054397045344674	M00960	PR
6	0.00610235387308533	M01267	FRA1
7	0.00622382607264738	M00034	P53
8	0.00658560112764628	M00641	HSF
9	0.00684369030265919	M00925	AP1
10	0.00725263462741377	M00490	BACH2
11	0.00815631694054531	M00109	CEBP
12	0.00935695228765299	M00472	FOXO4
13	0.00956357903590355	M00290	FREAC2
14	0.00966698641278729	M01482	NKX32
15	0.0104874553655619	M00495	BACH1
16	0.0112953153885734	M00199	AP1
17	0.0136948554436459	M00415	AREB6
18	0.0137521449321824	M01308	SOX4
19	0.0140105921176069	M00240	NKX25
20	0.0167940732274563	M01652	P53
21	0.0180978904688386	M00517	AP1
22	0.0188141243060463	M00174	AP1
23	0.0200938310247062	M00042	SOX5
24	0.0204159410474702	M00222	HAND1E47
25	0.0216246863781411	M00473	FOXO1
26	0.0222214921582313	M01016	SOX17
27	0.0224486592767784	M00115	TAXCREB
28	0.0255214257136045	M00634	GCM
29	0.0256942742283939	M00203	GATA
30	0.0270908041483611	M01181	NKX32
31	0.0291202082625213	M00485	NKX22
32	0.0294492295458247	M00123	MYCMAX
33	0.0302606921849399	M00277	LMO2COM
34	0.031917955259521	M00456	FAC1
35	0.0328628974514428	M00924	AP1
36	0.0351896910724482	M01173	CREBP1
37	0.0383817603164872	M01322	NKX26
38	0.038679909306794	M00322	MYCMAX
39	0.0389878096272638	M01216	FOXO1
40	0.0403291461385736	M001168	SREBP
41	0.0409182413443351	M00122	USF
42	0.0443286590979773	M01655	P53

43	0.0458002547828497	M00734	CIZ
44	0.0482655055441047	M001287	NEUROD

Table 5.9: TRAP predicted TFBS within the -137kb short enhancer region. The motifs with P≤0.05, therefore significant prediction of binding interaction are listed. Chondrocyte related AP-1 and SOX9 sites are highlighted (He *et al.* 2016). The SOX9 site was validated in Chapter 3.8. This site is more highly ranked in the short region compared to the full -137kb enhancer (Table 5.7).

Rank	P value	Matrix ID	TF
1	000943188754867874	M00454	MERF2
2	0.00183381239492753	M01437	HOXC12
3	0.00323825844094916	M00810	SRF
4	0.00453184161957132	M01378	HOXA11
5	0.00472370704523217	M01361	HOXC10
6	0.00523910985435194	M01329	HOXC11
7	0.00639554054229985	M00922	SRF
8	0.00667399893549392	M00185	SRF
9	0.00705546212825858	M01416	HOXC9
10	0.00841494917762742	M01434	HOXD11
11	0.00985873671197013	M01257	SRF
12	0.0114995205642501	M01250	ESF2
13	0.0137421018625407	M00211	PADS
14	0.01430627001459	M00201	CEBP
15	0.0148914770437449	M01380	HOXD12
16	0.0155780820350045	M00119	MAX
17	0.015994224540852	M01480	OBOX5
18	0.0161183698492418	M00377	PAX4
19	0.0166895803422783	M01484	PITX1
20	0.0179483754213986	M01304	SRF
21	0.0179597399520495	M00156	RORA1
22	0.0181604143543989	M01373	CCDX1
23	0.0183319934880481	M01070	CMAF
24	0.0186961531951948	M01342	CDP
25	0.0203463168366016	M00956	AR
26	0.0205324829795811	M01007	SRF
27	0.0213053266786003	M00220	SREBP
28	0.0224500806985822	M00960	PR
29	0.0240728751486352	M01237	TRF1
30	0.0242684442352505	M01352	NKX29
31	0.0251487342734887	M00146	HSF1
32	0.0256334046057471	M01410	IRX4
33	0.0284275546905568	M00707	TFIIA
34	0.0287491364218309	M00485	NKX22
35	0.0291822167235998	M00063	IRF2
36	0.0309273809623511	M00665	SP3
37	0.0325679158276536	M01450	OBOX1
38	0.0336506890360163	M00212	POLY_C

39	0.0339363305222814	M00964	PXR
40	0.0364018210196211	M00737	E2F1DP2
41	0.0366643626964319	M01425	HNF1B
42	0.0367858452950779	M00769	AML
43	0.0381344674203669	M01333	HDX
44	0.0381454587578272	M00118	MYCMAX
45	0.0409053325504842	M01145	OBOX6
46	0.0418438276239931	M01485	IRX3
47	0.0423281289920011	M00634	GCM
48	0.0428884747755864	M01449	CDX2
49	0.042993276455945	M00433	HMX1
50	0.0443679646200701	M01234	IPF1
51	0.0454411329036714	M00121	USF
52	0.0482158700727762	M00511	ERR1
53	0.0482402658836824	M00122	USF
54	0.048729229807334	M00317	LDSPOLYA
55	0.0493121821900077	M01132	SF1

Table 5.10: TRAP predicted TF interaction in -148kb full enhancer region. The motifs with P≤0.05, therefore significant prediction of binding interaction are listed. Limb development and chondrocyte related AML (Runx) and Pitx1 sites are highlighted.

Rank	P value	Matrix ID	TF
1	0.000318497155432085	M00454	MRF2
2	0.0022769808648937	M01378	HOXA11
3	0.0027583436115064	M01329	HOXC11
4	0.00386853624917793	M00377	PAX4
5	0.00405130836780143	M01361	HOXC10
6	0.00579633622871956	M01480	OBOX5
7	0.00640675923602274	M01484	PITX1
8	0.00689435516525672	M01434	HOXD11
9	0.0071940199193069	M00156	RORA1
10	0.00812906053328555	M01410	IRX4
11	0.00903209552745543	M01437	HOXC12
12	0.00979235617060126	M00119	MAX
13	0.0117509288018152	M00707	TFIIA
14	0.0133039660299070	M01237	TRF1
15	0.0135223699952858	M00220	SREBP1
16	0.0148436235583799	M01380	HOXD12
17	0.0157852061889879	M01450	OBOX1
18	0.0169308672694924	M01485	IRX3
19	0.0172761419977946	M00122	USF
20	0.0177704288195416	M00511	ERR1
21	0.0178232354864194	M01445	OBOX6
22	0.0183828873800084	M01425	HNF1B
23	0.0193618360738852	M00769	AML
24	0.0202529067059094	M01107	RUSH1A
25	0.0208602917945904	M01234	IPF1

26	0.0214022206096860	M01352	NKX29
27	0.0216430949047122	M01132	SF1
28	0.0219791865425958	M00118	MYCMAX
29	0.0236553852740911	M01318	IRX3
30	0.0241546706933410	M00737	E2FDP2
31	0.0243549180447044	M00485	NKX22
32	0.0243839503522254	M00317	LDSPOLYA
33	0.0247704102343558	M00433	HMX1
34	0.0262604038856323	M00729	CDX2
35	0.026370823495063	M01260	STAT1
36	0.0270284837732278	M00121	USF
37	0.0284130303176033	M01359	DOBOX4
38	0.0285756228465024	M00727	SF1
39	0.0287935859021345	M01208	FLI1
40	0.0294428196842479	M01416	HOXC9
41	0.0298408439866519	M01472	IRX5
42	0.0303766121074784	M01447	PITX2
43	0.0303796835307774	M01366	OTX1
44	0.0306595874562258	M01440	LHX8
45	0.0316958532755759	M01145	CMYC
46	0.0348629925580416	M00526	GCNF
47	0.0362619955639563	M01589	ERR2
48	0.0366209798677032	M00050	E2F
49	0.0411667676097563	M01372	NKX22
50	0.0412111179317466	M00236	ARNT
51	0.0417435943199939	M00736	E2F1DP1
52	0.0418790542314423	M00738	E2F4DP1
53	0.0424206812509408	M00251	XBP1
54	0.0429628852428795	M01364	OBOX2
55	0.0432960549541755	M00360	PAX3
56	0.0435482835189122	M00320	MTATA
57	0.0445684889134201	M01373	CDX1
58	0.0446181330245736	M01387	OTX2
59	0.0446918911014534	M00097	PAX6
60	0.0453088549703167	M00739	E2F4DP2
61	0.0455486221951887	M01466	OBOX3
62	0.0456525182510448	M00538	HTF
63	0.0470503563320529	M00772	IRF
64	0.0483384129519699	M00999	AIRE
65	0.0492470107499716	M00740	E2F1DP1RB

Table 5.11: TRAP predicted TF interaction within the -148kb_short region. The motifs with P≤0.05, therefore significant prediction of binding interaction are listed. Limb development and chondrocyte related AML (Runx) and Pitx1 sites are highlighted.

Rank	P value	Matrix ID	TF
1	0.00268402186717798	M01069	GZF1
2	0.00371987034413579	M00311	ATATA

3	0.00488290022015736	M00518	PPARA
4	0.00609238102305076	M00769	AML
5	0.00726266121975117	M00281	RFX1
6	0.00806881579727858	M01204	SPIB
7	0.0115342056261466	M00987	FOXP1
8	0.0120209130450558	M00292	FREAC4
9	0.013796166871901	M00103	CLOX
10	0.0144290580263504	M00405	MEF2
11	0.0144600536270681	M01661	HBP1
12	0.0146755242327166	M01196	CTF1
13	0.0153895788616608	M00767	FXR
14	0.0155467523429637	M00007	ELK1
15	0.0167352681071924	M01314	LHX61
16	0.0177129084735395	M00410	SOX9
17	0.0185612946398211	M00806	NF1
18	0.0194385487872916	M00722	COREBINDINGFACTOR
19	0.0201235204005591	M00731	OSF2
20	0.0216010754782094	M01422	LHX61
21	0.0226078060824153	M01431	BARX2
22	0.0230415057961467	M00102	CDP
23	0.0232133918913324	M00652	NRF1
24	0.0236724872074090	M01255	IPF1
25	0.0274242213582807	M00746	ELF1
26	0.030077354971388	M00105	CDPCR3
27	0.0315184760215368	M01216	FOXO1
28	0.0316509279154596	M01456	HMBOX1
29	0.0324925785430683	M00095	CDP
30	0.0389404604084379	M01150	DMRT5
31	0.0394546900307466	M01658	AML1
32	0.0434740805610291	M00478	CDC5
33	0.0436340785450632	M00211	PADS
34	0.044538478156510	M01081	ZEC
35	0.0446996657340436	M00236	ARNT
36	0.0448052116164019	M01362	CART1
37	0.0462422514780503	M01401	LBX2
38	0.0463091762197797	M01367	LHX9
39	0.0471099613571837	M00751	AML1
40	0.0477574887770392	M00203	GATA
41	0.0490929621456843	M00729	CDX2
42	0.0496305441822428	M00476	FOXO4

Table 5.12: TRAP predicted binding within the -230kb enhancer sequence. Chondrocyte related SOX9 and OSF2 (Runx2) are highlighted. The SOX9 site was validated in EMSA in Chapter 3.8

5.4.2 Optimisation of EMSA protein-probe binding conditions

Transcription factor-EMSA probe binding conditions were optimised using various additives in the Li-Cor Odyssey® EMSA Buffer kit as outlined in Chapter 2.9.2. Reactions that generated the strongest bands were used for further experiments

SOX9-probe interaction was optimised for the -137kb and -230kb wild-type probes as demonstrated in Figures 5.19 and 5.20.

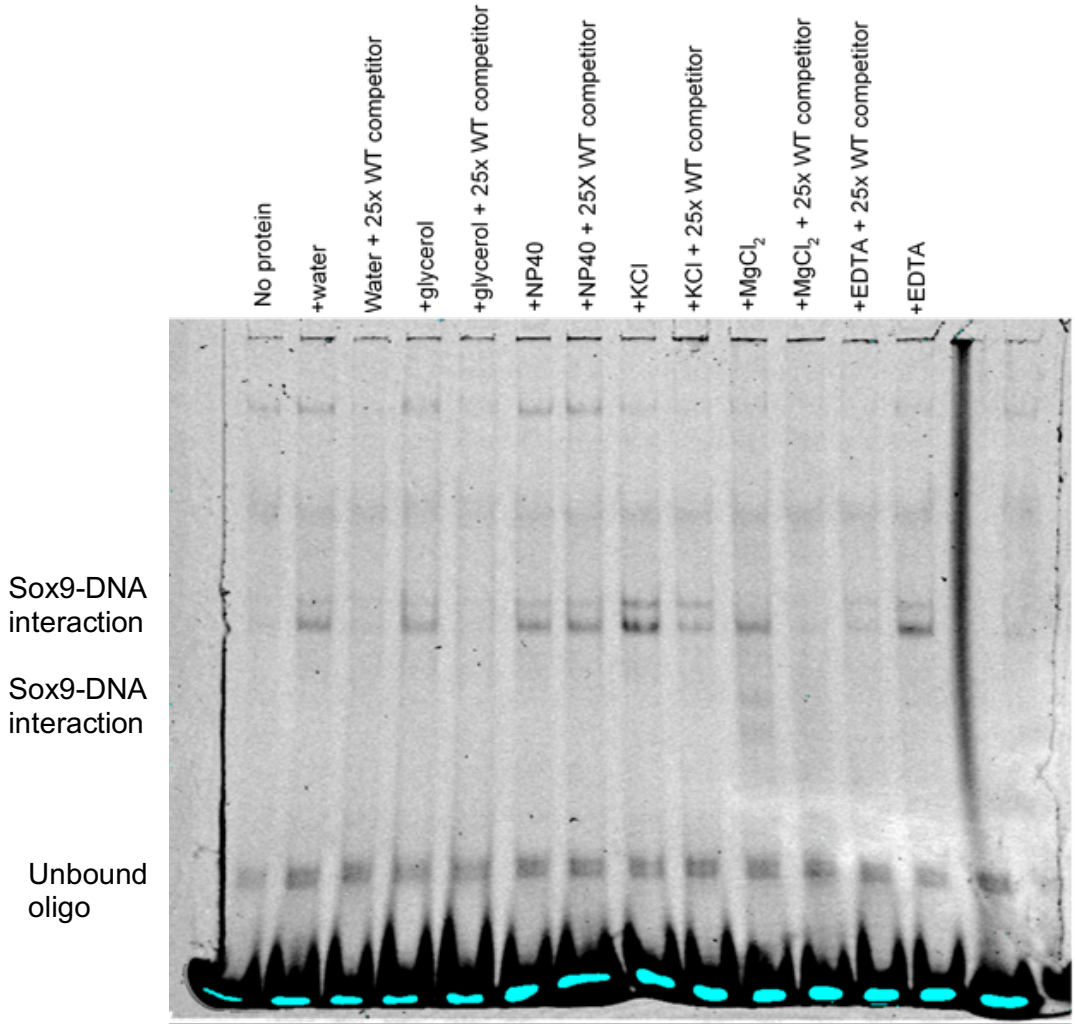


Figure 5.19: Optimisation of binding condition for EMSA of SOX9 binding in –137kb. Bands were present in each well where reaction with SOX9 protein was loaded, the intensity of bands was fainter in lanes where competitor oligos were added to reactions. KCl gave cleanest band and was chosen for subsequent EMSA exploring this protein-DNA interaction

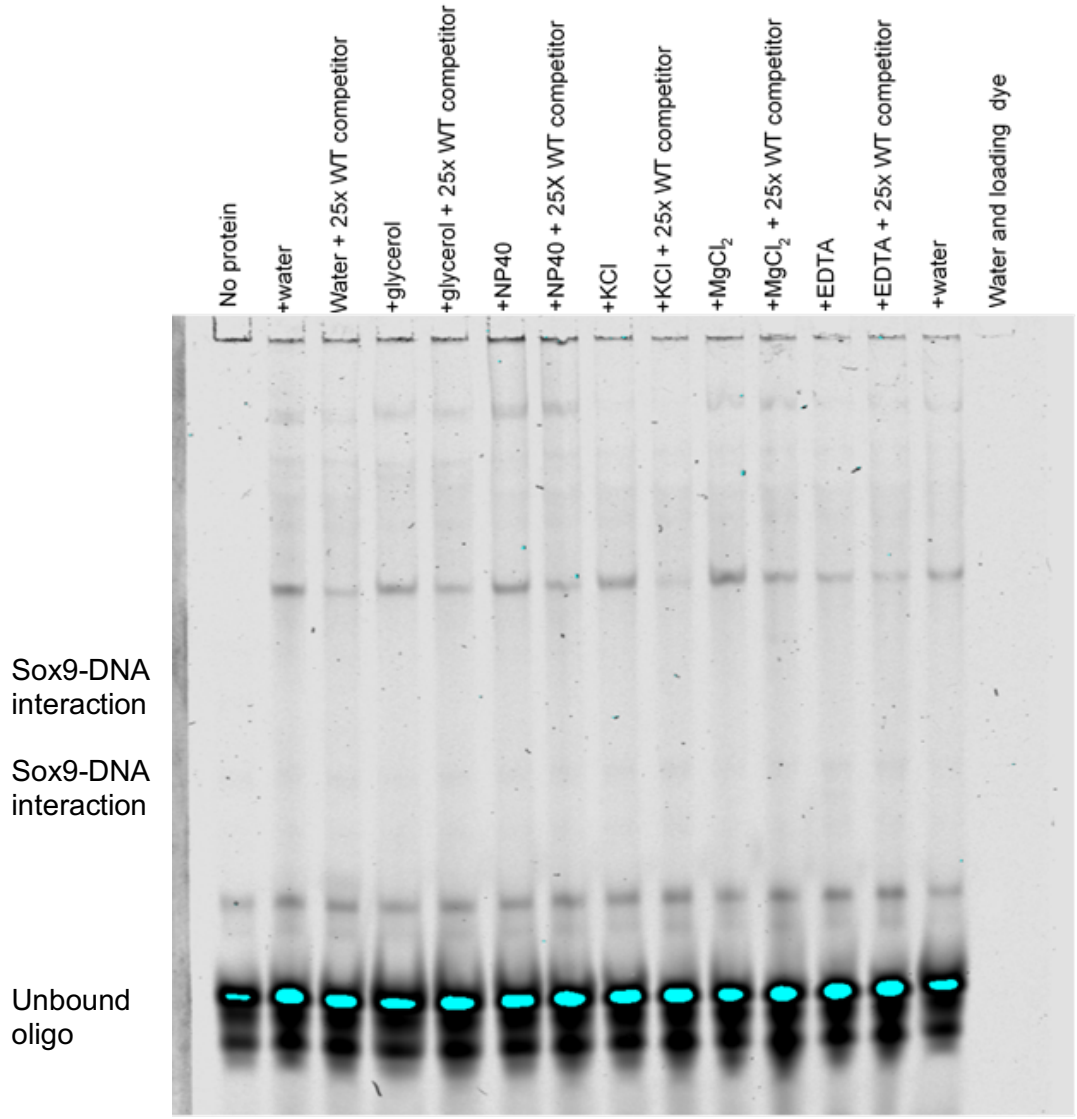


Figure 5.20: Optimisation of binding condition for EMSA of SOX9 binding in -230kb SOX9 EMSA probe. As with the -137kb probe, KCl was chosen as binding reaction additive as a clear band was observed denoting SOX9-probe interaction.

For the examination of Runx2 binding within -148kb, reaction conditions were optimised using a control probe that has previously been demonstrated to bind Runx2 (Lamour *et al.* 2007), as illustrated in Figure 5.21.

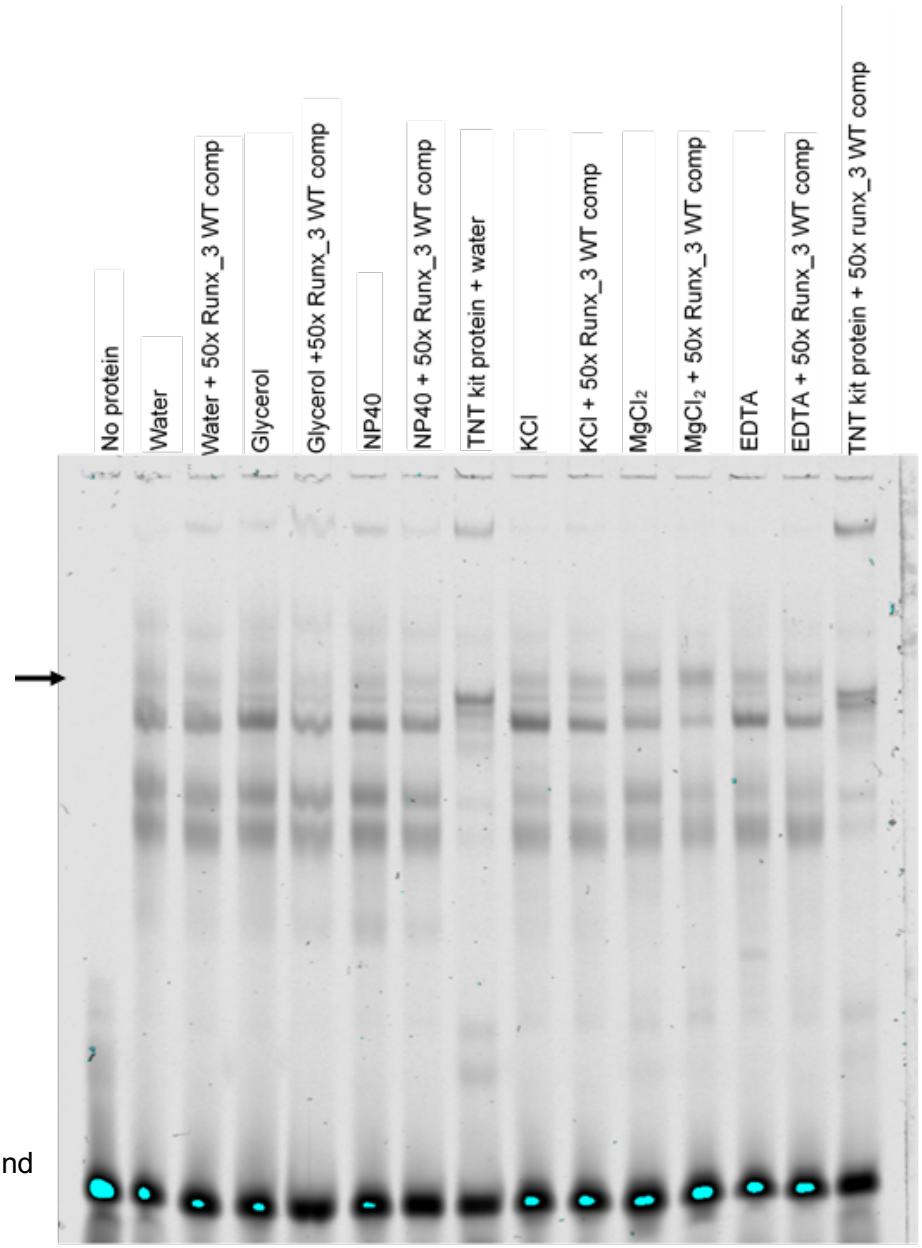


Figure 5.21: Optimisation of Runx2-EMSA probe interaction through the use of varying reaction additives. $MgCl_2$ was chosen as reaction additive as the banding pattern in this reaction most closely resembled that from the work of Lamour et al. (2007) (black arrowhead).

5.5 Further in silico resources in the prediction of CCN2 localisation within a TAD

The 3D- genome Interaction Viewer (3DIV) (Yang *et al.* 2018) was used to visualise interactive maps of chromatin interactions in the vicinity of human samples, as illustrated in Figures 5.22-23 which were lifted over into the mouse mm9 genome for predictions of the TAD in which *CCN2* is localised.

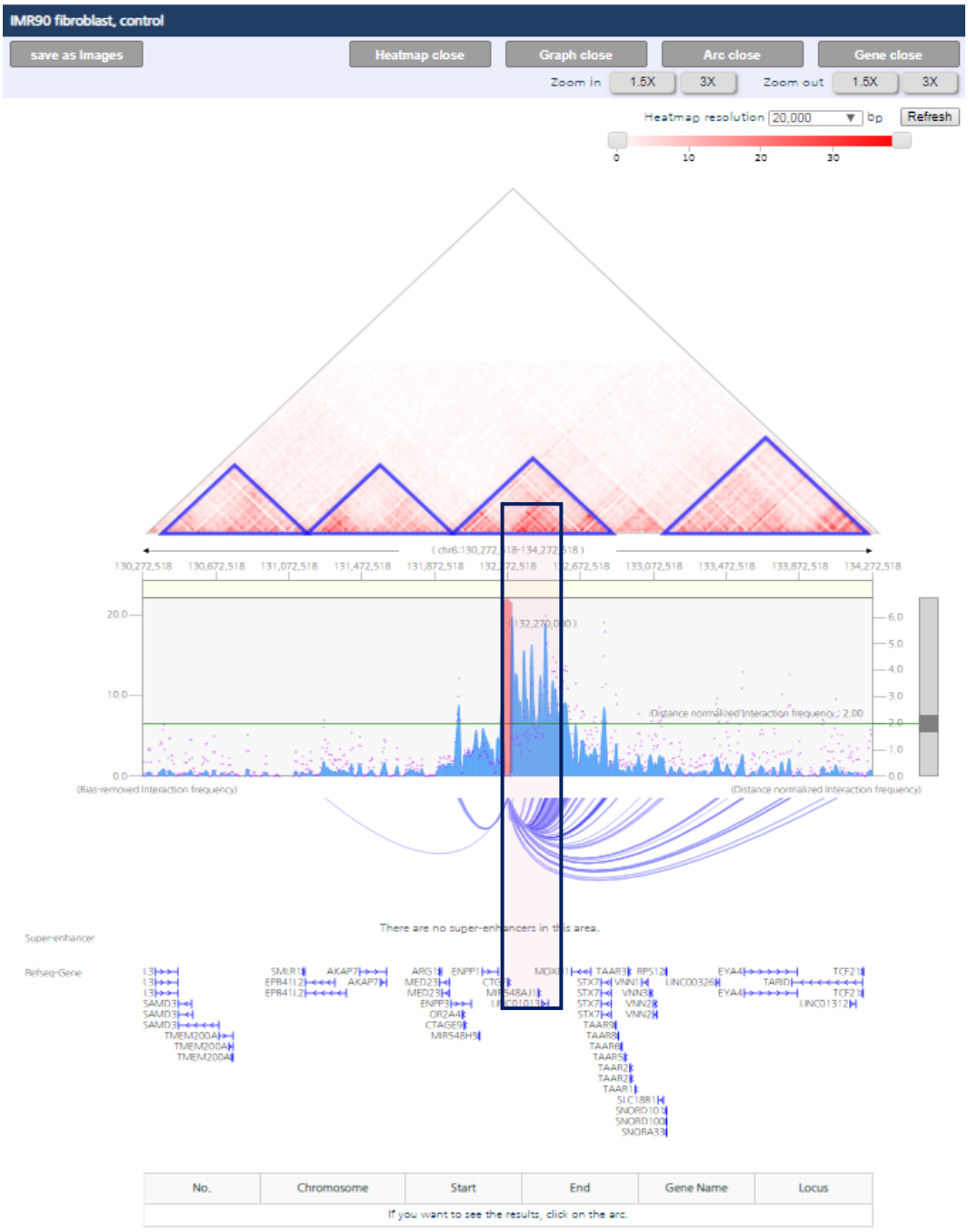


Figure 5.22: Screenshot of 3DIV IMR90 human fibroblast based chromatin interactions around *CCN2* (*CTGF*). Predicted TAD is highlighted with a black box. This region was lifted over into the mouse genome for further comparisons of TAD organisation.

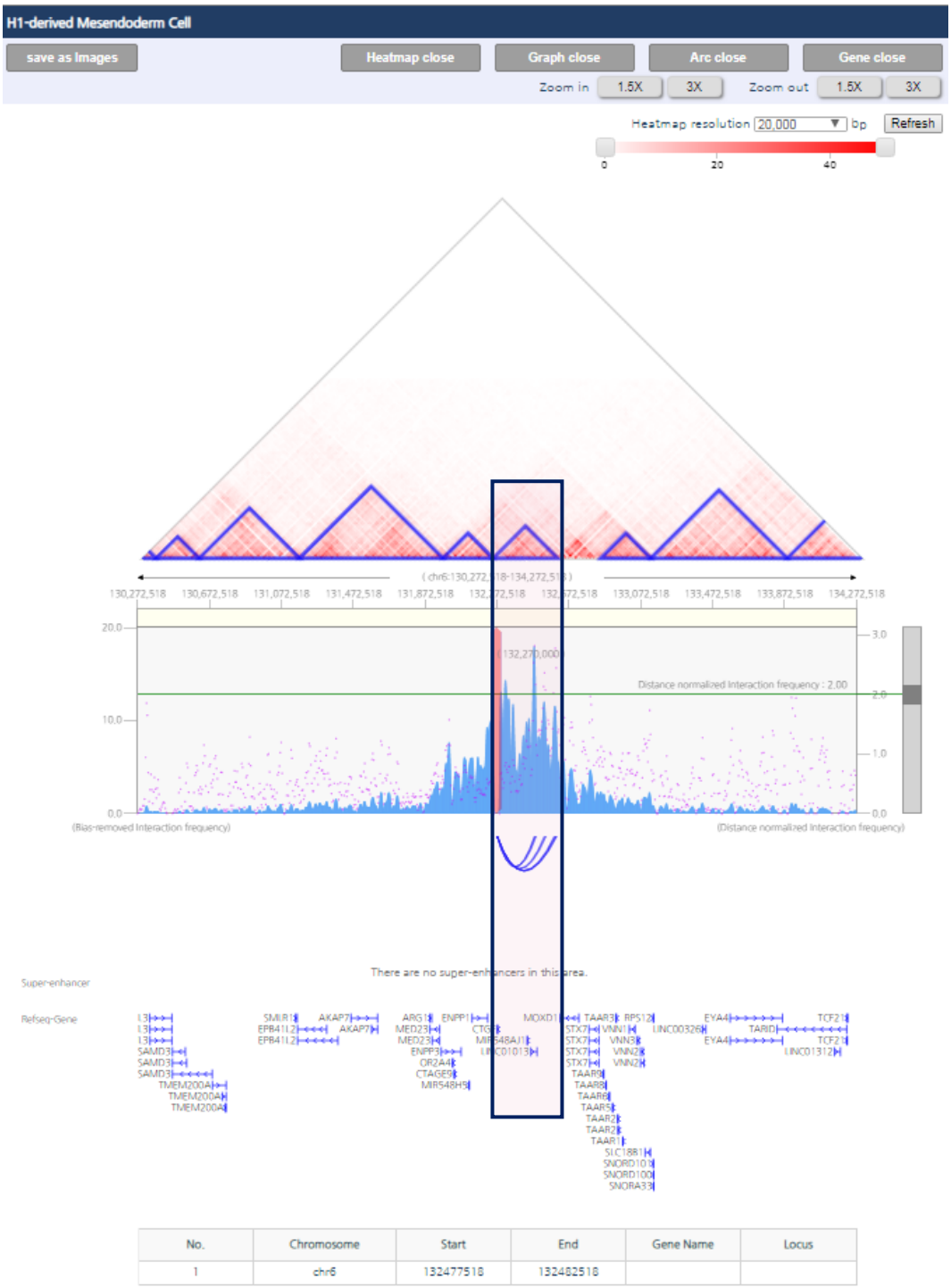


Figure 5.23: Screenshot of 3DIV h1-derived mesoderm human based chromatin interactions around *CCN2* (*CTGF*). Predicted TAD is highlighted with a black box. This region was lifted over into the mouse genome for further comparisons of TAD organisation.

ChIA-PET datasets available as part of the publicly available Yue Lab platform from http://promoter.bx.psu.edu/hi-c/chiapet.php (Wang *et al.* 2018) were also lifted over into the mm9 genome for TAD organisation These data included for CTCF binding within human K562 cell line (Figure 5.24).

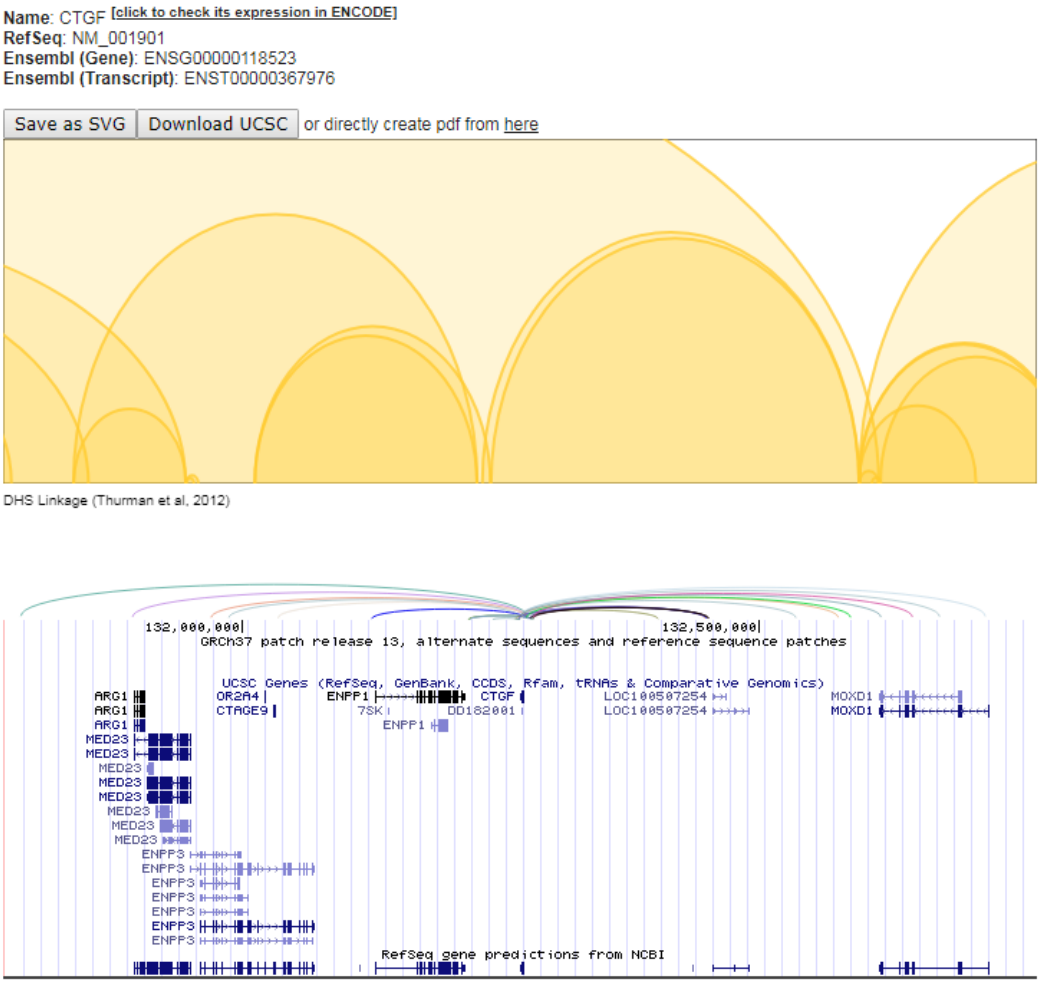


Figure 5.24: screenshot of publicly ChIA-PET data pertaining to CTCF interaction in human immortalised cell line (Wang *et al.* 2018)**.** Looping interactions between CTCF sites were visualised in the Yue Lab ChIA-PET resource available at: http://promoter.bx.psu.edu/hi-c/chiapet.php. These regions were lifted over into the mouse genome for TAD prediction.

This resource also contains Hi-ChIP datasets for cohesin interaction within mESC which was also used in the prediction of TAD organisation (Figure 5.25).

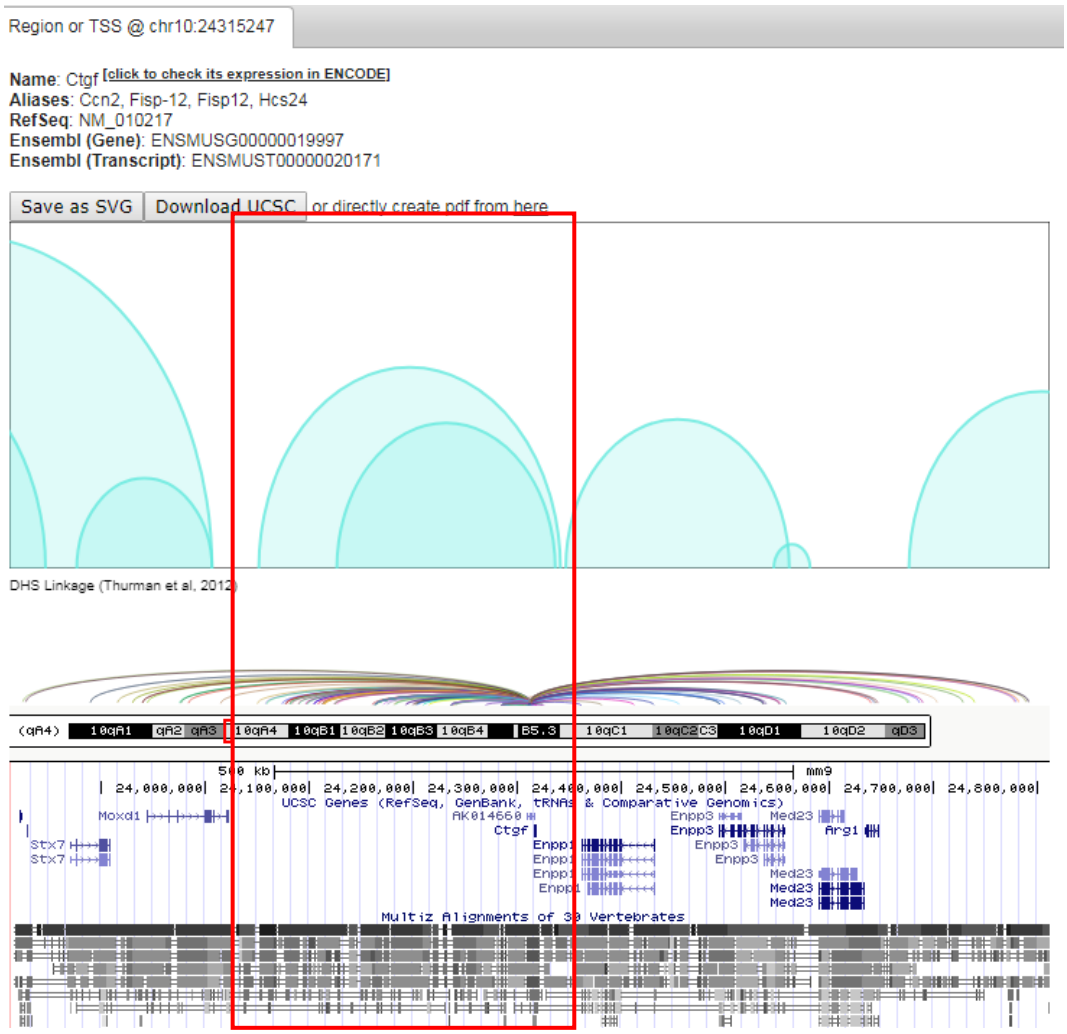


Figure 5.25: screenshot of Hi-ChIP generated publicly available data concerning cohesin interaction in mESC. The TAD region in which *CCN2* (*CTGF*) is predicted to sit is highlighted with a red box which stretches from close to the nearest 5' neighbour of *CCN2* within the mouse genome, *Moxd1* to downstream of *CCN2*, tallying with the localisation of the enhancers described in the current project.

6. References

Abd El Kader, T., Kubota, S., Nishida, T., … Takigawa, M. (2014). The regenerative effects of CCN2 independent modules on chondrocytes in vitro and osteoarthritis models in vivo. *Bone*, **59**, 180–188.

Abraham, D. J., Shiwen, X., Black, C. M., Sa, S., Xu, Y., & Leask, A. (2000). Tumor necrosis factor α suppresses the induction of connective tissue growth factor by transforming growth factor-β in normal and scleroderma fibroblasts. *Journal of Biological Chemistry*, **275**(20), 15220–15225.

Abreu, J. G., Ketpura, N. I., Reversade, B., & De Robertis, E. M. (2002). Connective-tissue growth factor (CTGF) modulates cell signalling by BMP and TGF-β. *Nature Cell Biology*, **4**(8), 599–604.

Aikawa, T., Gunn, J., Spong, S. M., Klaus, S. J., & Korc, M. (2006). Connective tissue growth factor-specific antibody attenuates tumor growth, metastasis, and angiogenesis in an orthotopic mouse model of pancreatic cancer. *Molecular Cancer Therapeutics*, **5**(5), 1108–1116.

Akiyama, H., Chaboissier, M. C., Martin, J. F., Schedl, A., & De Crombrugghe, B. (2002). The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes and Development*, **16**(21), 2813–2828.

Akiyama, H., Lyons, J. P., Mori-Akiyama, Y., … de Crombrugghe, B. (2004). Interactions between Sox9 and beta-catenin control chondrocyte differentiation. *Genes & Development*, **18**(9), 1072–87.

Akiyama, H., Shukunami, C., Nakamura, T., & Hiraki, Y. (2000). Differential Expressions of BMP Family Genes during Chondrogenic Differentiation of Mouse ATDC5 Cells. *Cell Structure and Function*, **25**(3), 195–204.

Andersson, R., Gebhard, C., Miguel-Escalada, I., … Sandelin, A. (2014). An atlas of active enhancers across human cell types and tissues. *Nature*, **507**(7493), 455–461.

Aoyama, E., Hattori, T., Hoshijima, M., … Takigawa, M. (2009). N-terminal domains of CCN family 2/connective tissue growth factor bind to aggrecan. *Biochemical Journal*, **420**, 413–420.

Aran, D., & Hellman, A. (2013). DNA methylation of transcriptional enhancers and cancer predisposition. *Cell*, **154**, 11–13.

Arner, E., Daub, C. O., Vitting-Seerup, K., … Beckhouse, A. (2015). Transcribed enhancers leadwaves of coordinated transcription in transitioning mammalian cells. *Science*, **347**(6225), 1010–1014.

Arnott, J. A., Lambi, A. G., Mundy, C., … Popoff, S. N. (2011). The role of connective tissue growth factor (CTGF/CCN2) in skeletogenesis. *Critical Reviews in Eukaryotic Gene Expression*, **21**(1), 43–69.

Arnott, J. A., Nuglozeh, E., Rico, M. C., … Popoff, S. N. (2007). Connective tissue growth factor (CTGF/CCN2) is a downstream mediator for TGF-β1-induced extracellular matrix production in osteoblasts. *Journal of Cellular Physiology*, **210**(3), 843–852.

Arnott, J. A., Zhang, X., Sanjay, A., … Popoff, S. N. (2008). Molecular requirements for induction of CTGF expression by TGF-β1 in primary osteoblasts. *Bone*, **42**(5), 871–885.

Ball, D. K., Rachfal, A. W., Kemper, S. A., & Brigstock, D. R. (2003). The heparin-binding 10 kDa fragment of connective tissue growth factor (CTGF) containing module 4 alone stimulates cell adhesion. *Journal of Endocrinology*, pp. 1–7.

Bar Oz, M., Kumar, A., Elayyan, J., … Dvir-Ginzberg, M. (2016). Acetylation reduces SOX9 nuclear entry and ACAN gene transactivation in human chondrocytes. *Aging Cell*, **15**(3),

499–508.

Barrett, T., Wilhite, S. E., Ledoux, P., … Soboleva, A. (2012). NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research*, **41**(D1), D991–D995.

Bartman, C. R., Hsu, S. C., Hsiung, C. C. S., Raj, A., & Blobel, G. A. (2016). Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Molecular Cell*, **62**(2), 237–247.

Barutcu, A. R., Tai, P. W. L., Wu, H., … Stein, G. S. (2014). The bone-specific Runx2-P1 promoter displays conserved three-dimensional chromatin structure with the syntenic Supt3h promoter. *Nucleic Acids Research*, **42**(16), 10360–10372.

Bernard, P., Tang, P., Liu, S., Dewing, P., Harley, V. R., & Vilain, E. (2003). Dimerization of SOX9 is required for chondrogenesis, but not for sex determination. *Human Molecular Genetics*, **12**(14), 1755–1765.

Bi, W., Deng, J. M., Zhang, Z., Behringer, R. R., & De Crombrugghe, B. (1999). Sox9 is required for cartilage formation. *Nature Genetics*, **22**(1), 85–89.

Bi, W., Huang, W., Whitworth, D. J., … de Crombrugghe, B. (2001). Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. *Proceedings of the National Academy of Sciences*, **98**(12), 6698–6703.

Blanchette, M., Kent, W. J., Riemer, C., … Miller, W. (2004). Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Research*, **14**(4), 708–15.

Bogdanović, O., Fernandez-Miñán, A., Tena, J. J., … Gómez-Skarmeta, J. L. (2012). Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Research*, **22**(10), 2043–2053.

Bonev, B., & Cavalli, G. (2016). Organization and function of the 3D genome. *Nature Reviews Genetics*, pp. 661–678.

Bonn, S., Zinzen, R. P., Girardot, C., … Furlong, E. E. M. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature Genetics*, **44**(2), 148–156.

Bork, P. (1993). The modular architecture of a new family of growth regulators related to connective tissue growth factor. *FEBS Letters*, pp. 125–130.

Bradham, D. M., Igarashi, A., Potter, R. L., & Grotendorst, G. R. (1991). Connective tissue growth factor: a cysteine-rich mitogen secreted by human vascular endothelial cells is related to the SRC-induced immediate early gene product CEF-10. *The Journal of Cell Biology*, **114**(6), 1285–94.

Busslinger, G. A., Stocsits, R. R., Van Der Lelij, P., … Peters, J. M. (2017). Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature*, **544**(7651), 503–507.

Calo, E., & Wysocka, J. (2013). Modification of Enhancer Chromatin: What, How, and Why? *Molecular Cell*, **49**(5), 825–837.

Canalis, E., Zanotti, S., & Smerdel-Ramoya, A. (2014). Connective tissue growth factor is a target of notch signaling in cells of the osteoblastic lineage. *Bone*, **64**, 273–280.

Cannavò, E., Khoueiry, P., Garfield, D. A., … Furlong, E. E. M. (2016). Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Current Biology*, **26**(1), 38–51.

Capellini, T. D., Chen, H., Cao, J., … Kingsley, D. M. (2017). Ancient selection for derived alleles at a GDF5 enhancer influencing human growth and osteoarthritis risk. *Nature Genetics*, **49**(8), 1202–1210.

Catarino, R. R., & Stark, A. (2018, February 1). Assessing sufficiency and necessity of

enhancer activities for gene expression and the mechanisms of transcription activation. *Genes and Development*, Cold Spring Harbor Laboratory Press, pp. 202–223.

Chen, D., Shen, J., Zhao, W., … Im, H.-J. (2017). Osteoarthritis: toward a comprehensive understanding of pathological mechanism. *Bone Research*, **5**(August 2016), 16044.

Chen, H., Capellini, T. D., Schoor, M., Mortlock, D. P., Reddi, A. H., & Kingsley, D. M. (2016). Heads, Shoulders, Elbows, Knees, and Toes: Modular Gdf5 Enhancers Control Different Joints in the Vertebrate Skeleton. *PLoS Genetics*, **12**(11).

Chen, L., Fink, T., Zhang, X. Y., Ebbesen, P., & Zachar, V. (2005). Quantitative transcriptional profiling of ATDC5 mouse progenitor cells during chondrogenesis. *Differentiation*, **73**(7), 350–363.

Chimal-Monroy, J., Rodriguez-Leon, J., Montero, J. A., … Hurle, J. M. (2003). Analysis of the molecular cascade responsible for mesodermal limb chondrogenesis: Sox genes and BMP signaling. *Developmental Biology*, **257**(2), 292–301.

Chiou, M. J., Chao, T. T., Wu, J. L., Kuo, C. M., & Chen, J. Y. (2006). The physiological role of CTGF/CCN2 in zebrafish notochond development and biological analysis of the proximal promoter region. *Biochemical and Biophysical Research Communications*, **349**(2), 750–758.

Cho, Y., Silverstein, R., Geisinger, M. T., … Arnott, J. A. (2015). AFAP1 is a novel downstream mediator of TGF-??1 for CCN2 induction in osteoblasts. *PLoS ONE*, **10**(9), 1–15.

Chronis, C., Fiziev, P., Papp, B., … Plath, K. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell*, **168**(3), 442–459.e20.

Cong, L., Ran, F. A., Cox, D., … Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, **339**, 819–823.

Core, L. J., Martins, A. L., Danko, C. G., Waters, C. T., Siepel, A., & Lis, J. T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature Genetics*, **46**(12), 1311–1320.

Core, L. J., Waterfall, J. J., Gilchrist, D. A., … Lis, J. T. (2012). Defining the Status of RNA Polymerase at Promoters. *Cell Reports*, **2**(4), 1025–1035.

Core, L. J., Waterfall, J. J., & Lis, J. T. (2008). Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science*, **322**, 1845–1848.

Coricor, G., Serra, R., Fox, A. J. S., … Wang, Y. (2016). TGF-β regulates phosphorylation and stabilization of Sox9 protein in chondrocytes through p38 and Smad dependent mechanisms. *Scientific Reports*, **6**, 38616.

Cotney, J., Leng, J., Oh, S., … Noonan, J. P. (2012). Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. *Genome Research*, **22**(6), 1069–1080.

Creyghton, M. P., Cheng, A. W., Welstead, G. G., … Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences*, **107**(50), 21931–21936.

Crick, F. (1970). Central dogma of molecular biology. *Nature*, **227**(5258), 561–563.

Daily, K., Patel, V. R., Rigor, P., Xie, X., & Baldi, P. (2011). MotifMap: Integrative genome-wide maps of regulatory motif sites for model species. *BMC Bioinformatics*, **12**(1), 495.

De Laat, W., & Duboule, D. (2013, October 23). Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature*, Nature Publishing Group, pp. 499–506.

de Santa, F., Barozzi, I., Mietton, F., … Natoli, G. (2010). A large fraction of extragenic RNA

Pol II transcription sites overlap enhancers. *PLoS Biology*, **8**(5), 1000384.

De Val, S., & Black, B. L. (2009). Transcriptional Control of Endothelial Cell Development. *Developmental Cell*, pp. 180–195.

Decker, R. S., Koyama, E., & Pacifici, M. (2014). Genesis and morphogenesis of limb synovial joints and articular cartilage. *Matrix Biology*, **39**, 5–10.

Decker, R. S., Koyama, E., & Pacifici, M. (2015). Articular Cartilage: Structural and Developmental Intricacies and Questions. *Current Osteoporosis Reports*, pp. 407–414.

Dickel, D. E., Visel, A., & Pennacchio, L. A. (2013). Functional anatomy of distant-acting mammalian enhancers. *Philosophical Transactions of the Royal Society B: Biological Sciences*.

Dickel, D. E., Ypsilanti, A. R., Pla, R., … Visel, A. (2018). Ultraconserved Enhancers Are Required for Normal Development. *Cell*, **172**(3), 491–499.e15.

Dixon, J. R., Gorkin, D. U., & Ren, B. (2016). Chromatin Domains: The Unit of Chromosome Organization. *Molecular Cell*, pp. 668–680.

Dixon, J. R., Selvaraj, S., Yue, F., … Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**(7398), 376–380.

Djebali, S., Davis, C. A., Merkel, A., … Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature*, **489**(7414), 101–108.

Dogan, N., Wu, W., Morrissey, C. S., … Hardison, R. C. (2015). Occupancy by key transcription factors is a more accurate predictor of enhancer activity than histone modifications or chromatin accessibility. *Epigenetics & Chromatin*, **8**(1), 16.

Donaghey, J., Thakurela, S., Charlton, J., … Meissner, A. (2018). Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nature Genetics*.

Dowen, J. M., Fan, Z. P., Hnisz, D., … Young, R. A. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell*, **159**(2), 374–387.

Dy, P., Wang, W., Bhattaram, P., … Lefebvre, V. (2012). Sox9 Directs Hypertrophic Maturation and Blocks Osteoblast Differentiation of Growth Plate Chondrocytes. *Developmental Cell*, **22**(3), 597–609.

Echtermeyer, F., Bertrand, J., Dreier, R., … Pap, T. (2009). Syndecan-4 regulates ADAMTS-5 activation and cartilage breakdown in osteoarthritis. *Nature Medicine*, **15**(9), 1072–1076.

Eguchi, T., Kubota, S., Kawata, K., … Takigawa, M. (2008). Novel transcription-factor-like function of human matrix metalloproteinase 3 regulating the CTGF/CCN2 gene. *Molecular and Cellular Biology*, **28**(7), 2391–413.

Eguchi, T., Kubota, S., Kondo, S., Kuboki, T., Yatani, H., & Takigawa, M. (2002). A novel cis-element that enhances connective tissue growth factor gene expression in chondrocytic cells. *Biochemical and Biophysical Research Communications*, **295**(2), 445–451.

Eguchi, T., Kubota, S., Kondo, S., … Takigawa, M. (2001). Regulatory mechanism of human connective tissue growth factor (CTGF/Hcs24) gene expression in a human chondrocytic cell line, HCS-2/8. *J Biochem (Tokyo)*, **130**(1), 79–87.

Eguchi, T., Kubota, S., & Takigawa, M. (2017). Promoter analyses of CCN genes. In *Methods in Molecular Biology*, Vol. 1489, Humana Press, New York, NY, pp. 177–185.

Ernst, J., Kheradpour, P., Mikkelsen, T. S., … Bernstein, B. E. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, **473**(7345), 43–49.

Ferrari, D., & Kosher, R. A. (2002). Dlx5 is a positive regulator of chondrocyte differentiation during endochondral ossification. *Developmental Biology*, **252**(2), 257–270.

Finger, F., Schörle, C., Zien, A., Gebhard, P., Goldring, M. B., & Aigner, T. (2003). Molecular phenotyping of human chondrocyte cell lines T/C-28a2, T/C-28a4, and C-28/I2. *Arthritis & Rheumatism*, **48**(12), 3395–3403.

Fleischer, T., Tekpli, X., Mathelier, A., … Kristensen, V. N. (2017). DNA methylation at enhancers identifies distinct breast cancer lineages. *Nature Communications*, **8**(1), 1379.

Fortin, J. P., & Hansen, K. D. (2015). Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. *Genome Biology*, **16**(1), 180.

Friedrichsen, S., Heuer, H., Christ, S., Cuthill, D., Bauer, K., & Raivich, G. (2005). Gene expression of connective tissue growth factor in adult mouse. *Growth Factors*, **23**(1), 43–53.

Friedrichsen, S., Heuer, H., Christ, S., … Raivich, G. (2003). CTGF expression during mouse embryonic development. *Cell and Tissue Research*, **312**(2), 175–188.

Frost, S. L., Liu, K., Li, I. M. H., … Bou-Gharios, G. (2018). Multiple enhancer regions govern the transcription of CCN2 during embryonic development. *Journal of Cell Communication and Signaling*, **12**(1), 231–243.

Fukaya, T., Lim, B., & Levine, M. (2016). Enhancer Control of Transcriptional Bursting. *Cell*, **166**(2), 358–368.

Fulco, C. P., Munschauer, M., Anyoha, R., … Engreitz, J. M. (2016). Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science*, **354**(6313), 769–773.

Gabay, O., Oppenheimer, H., Meir, H., Zaal, K., Sanchez, C., & Dvir-Ginzberg, M. (2012). Increased apoptotic chondrocytes in articular cartilage from adult heterozygous SirT1 mice. *Annals of the Rheumatic Diseases*, **71**(4), 613–616.

Garside, V. C., Zhao, Y., Marra, M. A., … Alder, O. (2015). SOX9 modulates the expression of key transcription factors required for heart valve development. *Development*, **142**(24), 4340–4350.

Gebauer, M., Saas, J., Sohler, F., … Liao, J. K. (2005). Comparison of the chondrosarcoma cell line SW1353 with primary human adult articular chondrocytes with regard to their gene expression profile and reactivity to IL-1beta. *Osteoarthritis and Cartilage / OARS, Osteoarthritis Research Society*, **13**(8), 697–708.

Geisinger, M. T., Astaiza, R., Butler, T., Popoff, S. N., Planey, S. L., & Arnott, J. A. (2012). Ets-1 is essential for connective tissue growth factor (CTGF/CCN2) induction by TGF-beta1 in osteoblasts. *PLoS ONE*, **7**(4).

Ghavi-Helm, Y., Klein, F. A., Pakozdi, T., … Furlong, E. E. M. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. *Nature*, **512**(1), 96–100.

Gibcus, J. H., & Dekker, J. (2013). The Hierarchy of the 3D Genome. *Molecular Cell*, pp. 773–782.

Gilchrist, D. A., Dos Santos, G., Fargo, D. C., … Adelman, K. (2010). Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell*, **143**(4), 540–551.

Goldring, M. B. (2012). Chondrogenesis, chondrocyte differentiation, and articular cartilage metabolism in health and osteoarthritis. *Therapeutic Advances in Musculoskeletal Disease*, **4**(4), 269–285.

Gong, S., Yu, H. H., Johnson, K. A., & Taylor, D. W. (2018a). DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Reports*, **22**(2), 359–371.

Gong, S., Zheng, C., Doughty, M. L., … Heintz, N. (2003). A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature*, **425**(6961), 917–925.

Gong, Y., Lazaris, C., Sakellaropoulos, T., … Tsirigos, A. (2018b). Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nature Communications*, **9**(1).

Graham, E., Moss, J., Burton, N., Armit, C., Richardson, L., & Baldock, R. (2015). The atlas of mouse development eHistology resource. *Development*, **142**(11), 1909–1911.

Gross, S., Krause, Y., Wuelling, M., & Vortkamp, A. (2012). Hoxa11 and hoxd11 regulate chondrocyte differentiation upstream of Runx2 and Shox2 in mice. *PLoS ONE*, **7**(8), e43553.

Grossman, S. R., Lander, E. S., Ray, J. P., Nguyen, T. H., Hacohen, N., & Engreitz, J. (2018). Positional specificity of different transcription factor classes within enhancers. *Proceedings of the National Academy of Sciences*, **115**(30), E7222–E7230.

Grotendorst, G. R. (2005). Individual domains of connective tissue growth factor regulate fibroblast proliferation and myofibroblast differentiation. *The FASEB Journal*, **19**(7), 729–738.

Grotendorst, G. R., Okochi, H., & Hayashi, N. (1996). A novel transforming growth factor beta response element controls the expression of the connective tissue growth factor gene. *Cell Growth & Differentiation*, **7**(4), 469–80.

Guo, X., Day, T. F., Jiang, X., Garrett-Beal, L., Topol, L., & Yang, Y. (2004). Wnt/β-catenin signaling is sufficient and necessary for synovial joint formation. *Genes and Development*, **18**(19), 2404–2417.

Guo, Y., Xu, Q., Canzio, D., Krainer, A. R., Maniatis, T., & Wu Correspondence, Q. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*, **162**, 900–910.

Haberle, V., & Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nature Reviews Molecular Cell Biology*, 1–17.

Hall-Glenn, F., Aivazi, A., Akopyan, L., … Lyons, K. M. (2013). CCN2/CTGF is required for matrix organization and to protect growth plate chondrocytes from cellular stress. *Journal of Cell Communication and Signaling*, **7**(3), 219–30.

Hall-Glenn, F., de Young, R. A., Huang, B. L., … Lyons, K. M. (2012). CCN2/Connective tissue growth factor is essential for pericyte adhesion and endothelial basement membrane formation during angiogenesis. *PLoS ONE*, **7**(2), e30562.

Hall-Glenn, F., & Lyons, K. M. (2011, October 20). Roles for CCN2 in normal physiological processes. *Cellular and Molecular Life Sciences*, SP Birkhäuser Verlag Basel, pp. 3209–3217.

Hall, B. K. (2015). Chapter 3 - Vertebrate Cartilages. In *Bones and Cartilage (Second Edition)*, pp. 43–59.

Han, Y., & Lefebvre, V. (2008). L-Sox5 and Sox6 drive expression of the aggrecan gene in cartilage by securing binding of Sox9 to a far-upstream enhancer. *Molecular and Cellular Biology*, **28**(16), 4999–5013.

Hanssen, L. L. P., Kassouf, M. T., Oudelaar, A. M., … Higgs, D. R. (2017). Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nature Cell Biology*, **19**(8), 952–961.

Harmston, N., & Lenhard, B. (2013). Chromatin and epigenetic features of long-range gene regulation. *Nucleic Acids Research*, **41**(15), 7185–7199.

Hartley, J. L., Temple, G. F., & Brasch, M. A. (2000). DNA cloning using in vitro site-specific

recombination. *Genome Research*, **10**(11), 1788–95.

Hattori, T., Muller, C., Gebhard, S., … von der Mark, K. (2010). SOX9 is a major negative regulator of cartilage vascularization, bone marrow formation and endochondral ossification. *Development*, **137**(6), 901–911.

Hay, D., Hughes, J. R., Babbs, C., … Higgs, D. R. (2016). Genetic dissection of the α-globin super-enhancer in vivo. *Nature Genetics*, **48**(8), 895–903.

He, X., Ohba, S., Hojo, H., & McMahon, A. P. (2016). AP-1 family members act with Sox9 to promote chondrocyte hypertrophy. *Development*, **143**(16), 3012–3023.

Heinegård, D. (2009). Proteoglycans and more - From molecules to biology. *International Journal of Experimental Pathology*, **90**(6), 575–586.

Heinegård, D., & Saxne, T. (2011). The role of the cartilage matrix in osteoarthritis. *Nature Reviews Rheumatology*, pp. 50–56.

Heintzman, N. D., Stuart, R. K., Hon, G., … Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*, **39**(3), 311–318.

Heinz, S., Romanoski, C. E., Benner, C., & Glass, C. K. (2015, February 4). The selection and function of cell type-specific enhancers. *Nature Reviews Molecular Cell Biology*, Nature Research, pp. 144–154.

Henriques, T., Scruggs, B. S., Inouye, M. O., … Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. *Genes and Development*, **32**(1), 26–41.

Henry, S. P., Liang, S., Akdemir, K. C., & De Crombrugghe, B. (2012). The postnatal role of Sox9 in cartilage. *Journal of Bone and Mineral Research*, **27**(12), 2511–2525.

Herz, J., & Strickland, D. K. (2001). LRP: a multifunctional scavenger and signaling receptor. *The Journal of Clinical Investigation*, **108**(6), 779–84.

Higgins, D. F., Biju, M. P., Akai, Y., Wutz, A., Johnson, R. S., & Haase, V. H. (2004). Hypoxic induction of Ctgf is directly mediated by Hif-1. *American Journal of Physiology. Renal Physiology*, **287**, F1223–F1232.

Hiyama, A., Morita, K., Sakai, D., & Watanabe, M. (2018). CCN family member 2/connective tissue growth factor (CCN2/CTGF) is regulated by Wnt-β-catenin signaling in nucleus pulposus cells. *Arthritis Research and Therapy*, **20**(1), 217.

Ho, L., & Crabtree, G. R. (2010). Chromatin remodelling during development. *Nature*, pp. 474–484.

Hojo, H., Ohba, S., He, X., Lai, L. P., & McMahon, A. P. (2016). Sp7/Osterix Is Restricted to Bone-Forming Vertebrates where It Acts as a Dlx Co-factor in Osteoblast Specification. *Developmental Cell*, **37**(3), 238–253.

Holbourn, K. P., Acharya, K. R., & Perbal, B. (2008). The CCN family of proteins: structure-function relationships. *Trends in Biochemical Sciences*, pp. 461–473.

Holmes, A., Abraham, D. J., Chen, Y., … Leask, A. (2003). Constitutive Connective Tissue Growth Factor Expression in Scleroderma Fibroblasts Is Dependent on Sp1. *Journal of Biological Chemistry*, **278**(43), 41728–41733.

Holmes, A., Abraham, D. J., Sa, S., Shiwen, X., Black, C. M., & Leask, A. (2001). CTGF and SMADs, Maintenance of Scleroderma Phenotype Is Independent of SMAD Signaling. *Journal of Biological Chemistry*, **276**(14), 10594–10601.

Horii, T., Arai, Y., Yamazaki, M., … Hatada, I. (2014). Validation of microinjection methods for generating knockout mice by CRISPR/Cas-mediated genome engineering. *Scientific Reports*, **4**.

Hoshijima, M., Hattori, T., Inoue, M., … Takigawa, M. (2006). CT domain of CCN2/CTGF directly interacts with fibronectin and enhances cell adhesion of chondrocytes through integrin α5β1. *FEBS Letters*, **580**(5), 1376–1382.

Hsu, P. D., Scott, D. A., Weinstein, J. A., … Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, **31**(9), 827–832.

Huang, B. L., Brugger, S. M., & Lyons, K. M. (2010). Stage-specific control of connective tissue growth factor (CTGF/CCN2) expression in chondrocytes by Sox9 and β-catenin. *Journal of Biological Chemistry*, **285**(36), 27702–27712.

Huang, J., Liu, X., Li, D., … Xu, J. (2016). Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. *Developmental Cell*, **36**(1), 9–23.

Huang, W., Chung, U. I., Kronenberg, H. M., … Crombrugghe, B. de. (2001). The chondrogenic transcription factor Sox9 is a target of signaling by the parathyroid hormone-related peptide in the growth plate of endochondral bones. *Proceedings of the National Academy of Sciences of the United States of America*, **98**(1), 160–165.

Hubbard, T., Barker, D., Birney, E., … Clamp, M. (2002). The Ensembl genome database project. *Nucleic Acids Research*, **30**(1), 38–41.

Hui, A. Y., McCarty, W. J., Masuda, K., Firestein, G. S., & Sah, R. L. (2012). A systems biology approach to synovial joint lubrication in health, injury, and disease. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, **4**(1), 15–37.

Infante, C. R., Mihala, A. G., Park, S., … Menke, D. B. (2015). Shared Enhancer Activity in the Limbs and Phallus and Functional Divergence of a Limb-Genital cis-Regulatory Element in Snakes. *Developmental Cell*, **35**(1), 107–119.

Itoh, S., Hattori, T., Tomita, N., … Takigawa, M. (2013). CCN Family Member 2/Connective Tissue Growth Factor (CCN2/CTGF) Has Anti-Aging Effects That Protect Articular Cartilage from Age-Related Degenerative Changes. *PLoS ONE*, **8**(8), e71156.

Ittner, L. M., & Götz, J. (2007). Pronuclear injection for the production of transgenic mice. *Nature Protocols*, **2**(5), 1206–1215.

Ivkovic, S., Yoon, B. S., Popoff, S. N., … Lyons, K. M. (2003). Connective tissue growth factor coordinates chondrogenesis and angiogenesis during skeletal development. *Development*, **130**(12), 2779–91.

Iwafuchi-Doi, M., & Zaret, K. S. (2014). Pioneer transcription factors in cell reprogramming. *Genes and Development*, pp. 2679–2692.

Jin, F., Li, Y., Dixon, J. R., … Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*, **503**(7475), 290–294.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, **337**, 816–820.

Jun, J. Il, & Lau, L. F. (2011). Taking aim at the extracellular matrix: CCN proteins as emerging therapeutic targets. *Nature Reviews Drug Discovery*, pp. 945–963.

Juven-Gershon, T., & Kadonaga, J. T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Developmental Biology*, pp. 225–229.

Kagey, M. H., Newman, J. J., Bilodeau, S., … Young, R. A. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, **467**(7314), 430–435.

Kaikkonen, M. U., Spann, N. J., Heinz, S., … Glass, C. K. (2013). Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Molecular Cell*, **51**(3), 310–325.

Kapoor, M., Liu, S., Huh, K., Parapuram, S., Kennedy, L., & Leask, A. (2008). Connective tissue growth factor promoter activity in normal and wounded skin. *Fibrogenesis & Tissue Repair*, **1**(1), 3.

Karsenty, G. (2008). Transcriptional Control of Skeletogenesis. *Annual Review of Genomics and Human Genetics*, **9**(1), 183–196.

Karsenty, G., & Wagner, E. F. (2002). Reaching a genetic and molecular understanding of skeletal development. *Developmental Cell*, pp. 389–406.

Kaufman, M. (2003). *The Atlas of Mouse Development*, 6th edn, London: Academic Press.

Kawaki, H., Kubota, S., Suzuki, A., … Takigawa, M. (2008). Functional requirement of CCN2 for intramembranous bone formation in embryonic mice. *Biochemical and Biophysical Research Communications*, **366**(2), 450–456.

Kawaki, H., Kubota, S., Suzuki, A., … Takigawa, M. (2011). Differential roles of CCN family proteins during osteoblast differentiation: Involvement of Smad and MAPK signaling pathways. *Bone*, **49**(5), 975–989.

Kawata, K., Eguchi, T., Kubota, S., … Takigawa, M. (2006). Possible role of LRP1, a CCN2 receptor, in chondrocytes. *Biochemical and Biophysical Research Communications*, **345**(2), 552–559.

Kent, W. J., Sugnet, C. W., Furey, T. S., … Haussler, D. (2002). The human genome browser at UCSC. *Genome Research*, **12**(6), 996–1006.

Khattab, H. M., Aoyama, E., Kubota, S., & Takigawa, M. (2015). Physical interaction of CCN2 with diverse growth factors involved in chondrocyte differentiation during endochondral ossification. *Journal of Cell Communication and Signaling*, **9**(3), 247–54.

Kieffer-Kwon, K. R., Tang, Z., Mathe, E., … Casellas, R. (2013). Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell*, **155**(7), 1507–1520.

Kim, H. S., Tan, Y., Ma, W., … Rosenfeld, M. G. (2018). Pluripotency factors functionally premark cell-type-restricted enhancers in ES cells. *Nature*, **556**(7702), 510–514.

Kim, T.-K., Hemberg, M., & Gray, J. M. (2015). Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers. *Cold Spring Harbor Perspectives in Biology*, **7**(1), a018622.

Kim, T. K., Hemberg, M., Gray, J. M., … Greenberg, M. E. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature*, **465**(7295), 182–187.

King, A. D., Huang, K., Rubbi, L., … Fan, G. (2016). Reversible Regulation of Promoter and Enhancer Histone Landscape by DNA Methylation in Mouse Embryonic Stem Cells. *Cell Reports*, **17**(1), 289–302.

Kobayashi, T., Chung, U. I., Schipani, E., … Kronenberg, H. M. (2002). PTHrP and Indian hedgehog control differentiation of growth plate chondrocytes at multiple steps. *Development (Cambridge, England)*, **129**(12), 2977–2986.

Kobayashi, T., Lyons, K. M., McMahon, A. P., & Kronenberg, H. M. (2005). BMP signaling stimulates cellular differentiation at multiple steps during cartilage development. *Proceedings of the National Academy of Sciences*, **102**(50), 18023–18027.

Komori, T., Yagi, H., Nomura, S., … Kishimoto, T. (1997). Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. *Cell*, **89**(5), 755–764.

Kondo, S., Kubota, S., Mukudai, Y., … Takigawa, M. (2006). Hypoxic regulation of stability of connective tissue growth factor/CCN2 mRNA by 3'-untranslated region interacting with a cellular protein in human chondrosarcoma cells. *Oncogene*, **25**(7), 1099–110.

Kothary, R., Clapoff, S., Darling, S., Perry, M. D., Moran, L. A., & Ant, J. R. (1989). Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. *Development*, **105**, 707–714.

Kouranova, E., Forbes, K., Zhao, G., … Cui, X. (2016). CRISPRs for Optimal Targeting: Delivery of CRISPR Components as DNA, RNA, and Protein into Cultured Cells and Single-Cell Embryos. *Human Gene Therapy*, **27**(6), 464–475.

Koyama, E., Shibukawa, Y., Nagayama, M., … Pacifici, M. (2008). A distinct cohort of progenitor cells participates in synovial joint and articular cartilage formation during mouse limb skeletogenesis. *Developmental Biology*, **316**(1), 62–73.

Koyama, E., Yasuda, T., Minugh-Purvis, N., … Pacifici, M. (2010). Hox11 genes establish synovial joint organization and phylogenetic characteristics in developing mouse zeugopod skeletal elements. *Development*, **137**(22), 3795–3800.

Kozhemyakina, E., Lassar, A. B., & Zelzer, E. (2015). A pathway to bone: signaling molecules and transcription factors involved in chondrocyte development and maturation. *Development*, **142**(5), 817–831.

Kronenberg, H. M. (2003). Developmental regulation of the growth plate. *Nature*, pp. 332–336.

Krupska, I., Bruford, E. A., & Chaqour, B. (2015). Eyeing the Cyr61/CTGF/NOV (CCN) group of genes in development and diseases: highlights of their structural likenesses and functional dissimilarities. *Human Genomics*, **9**, 24.

Kubota, S., Hattori, T., Nakanishi, T., & Takigawa, M. (1999). Involvement of cis-acting repressive element(s) in the 3'-untranslated region of human connective tissue growth factor gene. *FEBS Letters*, **450**(1–2), 84–88.

Kubota, S., Kondo, S., Eguchi, T., … Takigawa, M. (2000). Identification of an RNA element that confers post-transcriptional repression of connective tissue growth factor/hypertrophic chondrocyte specific 24 (ctgf/hcs24) gene: Similarities to retroviral RNA-protein interactions. *Oncogene*, **19**(41), 4773–4786.

Kubota, S., & Takigawa, M. (2011). The role of CCN2 in cartilage and bone development. *Journal of Cell Communication and Signaling*, **5**(3), 209–217.

Kulkarni, M. M., & Arnosti, D. N. (2003). Information display by transcriptional enhancers. *Development*, **130**(26), 6569–6575.

Kulyk, W. M., Rodgers, B. J., Greer, K., & Kosher, R. A. (1989). Promotion of embryonic chick limb cartilage differentiation by transforming growth factor-β. *Developmental Biology*, **135**(2), 424–430.

Kvon, E. Z. (2015). Using transgenic reporter assays to functionally characterize enhancers in animals. *Genomics*, pp. 185–192.

Kvon, E. Z., Kamneva, O. K., Melo, U. S., … Visel, A. (2016). Progressive Loss of Function in a Limb Enhancer during Snake Evolution. *Cell*, **167**(3), 633–642.e11.

Kwasnieski, J. C., Fiore, C., Chaudhari, H. G., & Cohen, B. A. (2014). High-throughput functional testing of ENCODE segmentation predictions. *Genome Research*, **24**(10), 1595–1602.

Labun, K., Montague, T. G., Gagnon, J. A., Thyme, S. B., & Valen, E. (2016). CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Research*, **44**(W1), W272–W276.

Lagha, M., Bothma, J. P., & Levine, M. (2012). Mechanisms of transcriptional precision in animal development. *Trends in Genetics*, **28**(8), 409–416.

Lam, M. T. Y., Cho, H., Lesch, H. P., … Glass, C. K. (2013). Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature*, **498**(7455), 511–515.

Lambi, A. G., Pankratz, T. L., Mundy, C., … Popoff, S. N. (2012). The Skeletal site-specific role of connective tissue growth factor in prenatal osteogenesis. *Developmental Dynamics*, **241**(12), 1944–1959.

Lamour, V., Detry, C., Sanchez, C., Henrotin, Y., Castronovo, V., & Bellahcène, A. (2007). Runx2- and histone deacetylase 3-mediated repression is relieved in differentiating human osteoblast cells to allow high bone sialoprotein expression. *The Journal of Biological Chemistry*, **282**(50), 36240–9.

Lanctôt, C., Moreau, A., Chamberland, M., Michel L., T., & Drouin, J. (1999). Hindlimb patterning and mandible development require the Ptx1 gene. *Development*, **125**(3), 1805–1810.

Leask, A., & Abraham, D. J. (2006). All in the CCN family: essential matricellular signaling modulators emerge from the bunker. *Journal of Cell Science*, **119**(23), 4803–4810.

Leask, A., Holmes, A., Black, C. M., & Abraham, D. J. (2003). Connective tissue growth factor gene regulation. Requirements for its induction by transforming growth factor-beta 2 in fibroblasts. *The Journal of Biological Chemistry*, **278**(15), 13008–15.

Leask, A., Parapuram, S. K., Shi-Wen, X., & Abraham, D. J. (2009). Connective tissue growth factor (CTGF, CCN2) gene regulation: A potent clinical bio-marker of fibroproliferative disease? *Journal of Cell Communication and Signaling*, **3**(2), 89–94.

LeBlanc, K. T., Walcott, M. E., Gaur, T., … Fanning, P. J. (2015). Runx1 Activities in Superficial Zone Chondrocytes, Osteoarthritic Chondrocyte Clones and Response to Mechanical Loading. *Journal of Cellular Physiology*, **230**(2), 440–448.

Lefebvre, V., & Dvir-Ginzberg, M. (2017). SOX9 and the many facets of its regulation in the chondrocyte lineage. *Connective Tissue Research*, **58**(1), 2–14.

Lefebvre, V., Li, P., & De Crombrugghe, B. (1998). A new long form of Sox5 (L-Sox5), Sox6 and Sox9 are coexpressed in chondrogenesis and cooperatively activate the type II collagen gene. *EMBO Journal*, **17**(19), 5718–5733.

Lefebvre, V., & Smits, P. (2005). Transcriptional Control of Chondrocyte Fate and Differentiation. In *Birth Defects Research Part C: Embryo Today: Reviews*, Vol. 75, John Wiley & Sons, Ltd, pp. 200–212.

Leonard, C. M., Fuld, H. M., Frenz, D. A., Downie, S. A., Massague, J., & Newman, S. A. (1991). Role of transforming growth factor-β in chondrogenic pattern formation in the embryonic limb: Stimulation of mesenchymal condensation and fibronectin gene expression by exogenenous TGF-β and evidence for endogenous TGF-β-like activity. *Developmental Biology*, **145**(1), 99–109.

Lettice, L. A., Heaney, S. J. H., Purdie, L. A., … de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics*, **12**(14), 1725–1735.

Levine, M., Cattoglio, C., & Tjian, R. (2014). Looping back to leap forward: Transcription enters a new era. *Cell*, pp. 13–25.

Li, G., Ruan, X., Auerbach, R. K., … Ruan, Y. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell*, **148**(1–2), 84–98.

Li, I. M. H., Liu, K., Neal, A., Clegg, P. D., De Val, S., & Bou-Gharios, G. (2018). Differential tissue specific, temporal and spatial expression patterns of the Aggrecan gene is modulated by independent enhancer elements. *Scientific Reports*, **8**(1), 950.

Li, S. W., Prockop, D. J., Helminen, H., … Khillan, J. S. (1995). Transgenic mice with targeted inactivation of the Col2a1 gene for collagen II develop a skeleton with membranous and periosteal bone but no endochondral bone. *Genes and Development*, **9**(22), 2821–

2830.

Li, W., Notani, D., Ma, Q., … Rosenfeld, M. G. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*, **498**(7455), 516–520.

Li, W., Notani, D., & Rosenfeld, M. G. (2016). Enhancers as non-coding RNA transcription units: Recent insights and future perspectives. *Nature Reviews Genetics*, pp. 207–223.

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., … Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**(5950), 289–293.

Little, C. B., Barai, A., Burkhardt, D., … Thompson, E. W. (2009). Matrix metalloproteinase 13-deficient mice are resistant to osteoarthritic cartilage erosion but not chondrocyte hypertrophy or osteophyte development. *Arthritis and Rheumatism*, **60**(12), 3723–3733.

Liu, C.-F., Angelozzi, M., Haseeb, A., & Lefebvre, V. (2018a). SOX9 is dispensable for the initiation of epigenetic remodeling and the activation of marker genes at the onset of chondrogenesis. *Development*, **145**(14), dev164459.

Liu, C.-F., & Lefebvre, V. (2015). The transcription factors SOX9 and SOX5/SOX6 cooperate genome-wide through super-enhancers to drive chondrogenesis. *Nucleic Acids Research*, **43**(17), 8183–203.

Liu, C.-F., Samsa, W. E., Zhou, G., & Lefebvre, V. (2017). Transcriptional control of chondrocyte specification and differentiation. *Seminars in Cell and Developmental Biology*, **62**, 34–49.

Liu, W., Ma, Q., Wong, K., … Rosenfeld, M. G. (2013). Brd4 and JMJD6-Associated Anti-Pause Enhancers in Regulation of Transcriptional Pause Release. *Cell*, **155**, 1581–1595.

Liu, Y., Chang, J.-C., Hon, C.-C., … Minoda, A. (2018b). Chromatin accessibility landscape of articular knee cartilage reveals aberrant enhancer regulation in osteoarthritis. *Scientific Reports*, **8**(1), 15499.

Long, F., & Ornitz, D. M. (2013, January 1). Development of the endochondral skeleton. *Cold Spring Harbor Perspectives in Biology*, Cold Spring Harbor Laboratory Press, p. a008334.

Long, H. K., Prescott, S. L., & Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell*, pp. 1170–1187.

Lopes, R., Korkmaz, G., & Agami, R. (2016). Applying CRISPR-Cas9 tools to identify and characterize transcriptional enhancers. *Nature Reviews Molecular Cell Biology*, pp. 597–604.

Lories, R. J., & Luyten, F. P. (2011). The bone-cartilage unit in osteoarthritis. *Nature Reviews Rheumatology*, pp. 43–49.

Luo, Q., Kang, Q., Si, W., … He, T.-C. C. (2004). Connective tissue growth factor (CTGF) is regulated by Wnt and bone morphogenetic proteins signaling in osteoblast differentiation of mesenchymal stem cells. *The Journal of Biological Chemistry*, **279**(53), 55958–68.

Luo, Y., Sinkeviciute, D., He, Y., … Bay-Jensen, A. (2017). The minor collagens in articular cartilage. *Protein and Cell*, **8**(8), 560–572.

Lupiáñez, D. G., Kraft, K., Heinrich, V., … Mundlos, S. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, **161**(5), 1012–1025.

Mackie, E. J., Ahmed, Y. A., Tatarczuch, L., Chen, K.-S. S., & Mirams, M. (2008). Endochondral ossification: how cartilage is converted into bone in the developing skeleton. *The International Journal of Biochemistry & Cell Biology*, **40**(1), 46–62.

Maeda-Uematsu, A., Kubota, S., Kawaki, H., … Takigawa, M. (2014). CCN2 as a novel

molecule supporting energy metabolism of chondrocytes. *Journal of Cellular Biochemistry*, **115**(5), 854–865.

Maeda, A., Nishida, T., Aoyama, E., … Takigawa, M. (2009). CCN family 2/connective tissue growth factor modulates BMP signalling as a signal conductor, which action regulates the proliferation and differentiation of chondrocytes. *Journal of Biochemistry*, **145**(2), 207–216.

Maes, C., Kobayashi, T., Selig, M. K., … Kronenberg, H. M. (2010). Osteoblast precursors, but not mature osteoblasts, move into developing and fractured bones along with invading blood vessels. *Developmental Cell*, **19**(2), 329–344.

Maes, C., & Kronenberg, H. M. (2016). Bone Development and Remodeling. In *Endocrinology: Adult and Pediatric*, W.B. Saunders, p. 1038--1062.e8.

Manke, T., Roider, H. G., & Vingron, M. (2008). Statistical modeling of transcription factor binding affinities predicts regulatory interactions. *PLoS Computational Biology*, **4**(3), 1000039.

Margueron, R., & Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature*, **469**(7330), 343–349.

Massagué, J., Seoane, J., & Wotton, D. (2005). Smad transcription factors. *Genes & Development*, **19**(23), 2783–810.

Maurano, M. T., Humbert, R., Rynes, E., … Stamatoyannopoulos, J. A. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science*, **337**(6099), 1190–1195.

Meng, H., & Bartholomew, B. (2018). Emerging roles of transcriptional enhancers in chromatin looping and promoter-proximal pausing of RNA polymerase II. *Journal of Biological Chemistry*, pp. 13786–13794.

Meyer, M. B., Benkusky, N. A., & Pike, J. W. (2014). The RUNX2 cistrome in osteoblasts: Characterization, down-regulation following differentiation, and relationship to gene expression. *Journal of Biological Chemistry*, **289**(23), 16016–16031.

Mikhaylichenko, O., Bondarenko, V., Harnett, D., … Furlong, E. E. M. (2018). The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes and Development*, **32**(1), 42–57.

Minina, E., Wenzel, H. M., Kreschel, C., … Vortkamp, A. (2001, October 15). BMP and Ihh/PTHrP signaling interact to coordinate chondrocyte proliferation and differentiation. *Development*, The Company of Biologists Ltd, pp. 4523–4534.

Mizuhashi, K., Ono, W., Matsushita, Y., … Ono, N. (2018). Resting zone of the growth plate houses a unique class of skeletal stem cells. *Nature*, **563**(7730), 254–258.

Mobasheri, A., Rayman, M. P., Gualillo, O., Sellam, J., Van Der Kraan, P., & Fearon, U. (2017). The role of metabolism in the pathogenesis of osteoarthritis. *Nature Reviews Rheumatology*, pp. 302–311.

Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M., & Valen, E. (2014). CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Research*, **42**(W1), W401–W407.

Moreno-Mateos, M. A., Vejnar, C. E., Beaudoin, J.-D., … Giraldez, A. J. (2015). CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nature Methods*, **12**(10), 982–988.

Mori, T., Kawara, S., Shinozaki, M., … Takehara, K. (1999). Role and interaction of connective tissue growth factor with transforming growth factor-beta in persistent fibrosis: A mouse fibrosis model. *Journal of Cellular Physiology*, **181**(1), 153–159.

Morikawa, M., Koinuma, D., Miyazono, K., & Heldin, C.-H. (2013). Genome-wide

mechanisms of Smad binding. *Oncogene*, **32**(10), 1609–1615.

Mousavi, K., Zare, H., Dell'Orso, S., … Sartorelli, V. (2013). ERNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci. *Molecular Cell*, **51**(5), 606–617.

Murakawa, Y., Yoshihara, M., Kawaji, H., … Hayashizaki, Y. (2016). Enhanced Identification of Transcriptional Enhancers Provides Mechanistic Insights into Diseases. *Trends in Genetics*, **32**(2), 76–88.

Murase, Y., Hattori, T., Aoyama, E., … Kubota, S. (2016). Role of CCN2 in Amino Acid Metabolism of Chondrocytes. *Journal of Cellular Biochemistry*, **117**(4), 927–937.

Muse, G. W., Gilchrist, D. A., Nechaev, S., … Adelman, K. (2007). RNA polymerase is poised for activation across the genome. *Nature Genetics*, **39**(12), 1507–1511.

Nakanishi, T., Nishida, T., Shimo, T., … Takigawa, M. (2000). Effects of CTGF/Hcs24, a product of a hypertrophic chondrocyte-specific gene, on the proliferation and differentiation of chondrocytes in culture. *Endocrinology*, **141**(1), 264–273.

Nakanishi, T., Yamaai, T., Asano, M., … Takigawa, M. (2001). Overexpression of connective tissue growth factor/hypertrophic chondrocyte-specific gene product 24 decreases bone density in adult mice and induces dwarfism. *Biochemical and Biophysical Research Communications*, **281**(3), 678–681.

Nakashima, K., Zhou, X., Kunkel, G., … de Crombrugghe, B. (2002). The Novel Zinc Finger-Containing Transcription Factor Osterix Is Required for Osteoblast Differentiation and Bone Formation. *Cell*, **108**(1), 17–29.

Nakata, E., Nakanishi, T., Kawai, A., … Takigawa, M. (2002). Expression of connective tissue growth factor/hypertrophic chondrocyte-specific gene product 24 (CTGF/Hcs24) during fracture healing. *Bone*, **31**(4), 441–447.

Natoli, G., & Andrau, J.-C. (2012). Noncoding transcription at enhancers: general principles and functional models. *Annu. Rev. Genet.*, **46**, 1–19.

Nishida, T., Emura, K., Kubota, S., Lyons, K. M., & Takigawa, M. (2011). CCN family 2/connective tissue growth factor (CCN2/CTGF) promotes osteoclastogenesis via induction of and interaction with dendritic cell-specific transmembrane protein (DC-STAMP). *Journal of Bone and Mineral Research*, **26**(2), 351–363.

Nishida, T., Kubota, S., Fukunaga, T., … Takigawa, M. (2003). CTGF/Hcs24, hypertrophic chondrocyte-specific gene product, interacts with perlecan in regulating the proliferation and differentiation of chondrocytes. *Journal of Cellular Physiology*, **196**(2), 265–275.

Nishida, T., Kubota, S., Kojima, S., … Takigawa, M. (2004). Regeneration of defects in articular cartilage in rat knee joints by CCN2 (connective tissue growth factor). *Journal of Bone and Mineral Research*, **19**(8), 1308–1319.

Nora, E. P., Goloborodko, A., Valton, A. L., … Bruneau, B. G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*, **169**(5), 930–944.e22.

Nord, A. S. (2015). Learning about mammalian gene regulation from functional enhancer assays in the mouse. *Genomics*, pp. 178–184.

Nord, A. S., Blow, M. J., Attanasio, C., … Visel, A. (2013). Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell*, **155**(7), 1521–1531.

Nowak, C. M., Lawson, S., Zerez, M., & Bleris, L. (2016). Guide RNA engineering for versatile Cas9 functionality. *Nucleic Acids Research*, **44**(10), 9555–9564.

Nozawa, K., Fujishiro, M., Kawasaki, M., … Sekigawa, I. (2009). Connective tissue growth factor promotes articular damage by increased osteoclastogenesis in patients with rheumatoid arthritis. *Arthritis Research and Therapy*, **11**(6), R174.

Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N., & Mirny, L. A. (2018). Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proceedings of the National Academy of Sciences*, **115**(29), E6697–E6706.

Oh, C., Yasuda, H., Zhao, W., … Chen, D. (2016). SOX9 directly Regulates CTGF/CCN2 Transcription in Growth Plate Chondrocytes and in Nucleus Pulposus Cells of Intervertebral Disc. *Scientific Reports*, **6**(1), 29916.

Oh, C. do, Maity, S. N., Lu, J. F., … Yasuda, H. (2010). Identification of SOX9 interaction sites in the genome of chondrocytes. *PLoS ONE*, **5**(4), e10113.

Ohba, S., He, X., Hojo, H., & McMahon, A. P. (2015). Distinct Transcriptional Programs Underlie Sox9 Regulation of the Mammalian Chondrocyte. *Cell Reports*, **12**(2), 229–243.

Omoto, S., Nishida, K., Yamaai, Y., … Takigawa, M. (2004). Expression and localization of connective tissue growth factor (CTGF/Hcs24/CCN2) in osteoarthritic cartilage. *Osteoarthritis and Cartilage*, **12**(10), 771–778.

Ong, C. T., & Corces, V. G. (2014). CTCF: An architectural protein bridging genome topology and function. *Nature Reviews Genetics*, pp. 234–246.

Ono, N., Ono, W., Nagasawa, T., & Kronenberg, H. M. (2014). A subset of chondrogenic cells provides early mesenchymal progenitors in growing bones. *Nature Cell Biology*, **16**(12), 1157–1167.

Ortega, N., Behonick, D. J., & Werb, Z. (2004, February 1). Matrix remodeling during endochondral ossification. *Trends in Cell Biology*, Elsevier Current Trends, pp. 86–93.

Osterwalder, M., Barozzi, I., Tissiéres, V., … Pennacchio, L. A. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature*, **554**(7691), 239–243.

Ostuni, R., Piccolo, V., Barozzi, I., … Natoli, G. (2013). Latent enhancers activated by stimulation in differentiated cells. *Cell*, **152**(1–2), 157–171.

Pachkov, M., Balwierz, P. J., Arnold, P., Ozonov, E., & Van Nimwegen, E. (2013). SwissRegulon, a database of genome-wide annotations of regulatory sites: Recent updates. *Nucleic Acids Research*, **41**(D1).

Papenbrock, T., Visconti, R. P., & Awgulewitsch, A. (2000). Loss of fibula in mice overexpressing Hoxc11. *Mechanisms of Development*, **92**(2), 113–123.

Parisi, M. S., Gazzerro, E., Rydziel, S., & Canalis, E. (2006). Expression and regulation of CCN genes in murine osteoblasts. *Bone*, **38**(5), 671–677.

Pennacchio, L. A., Ahituv, N., Moses, A. M., … Rubin, E. M. (2006). In vivo enhancer analysis of human conserved non-coding sequences. *Nature*, **444**(7118), 499–502.

Perbal, B. (2004). CCN proteins: Multifunctional signalling regulators. *Lancet*, pp. 62–64.

Perbal, B., Tweedie, S., & Bruford, E. (2018). The official unified nomenclature adopted by the HGNC calls for the use of the acronyms, CCN1–6, and discontinuation in the use of CYR61, CTGF, NOV and WISP 1–3 respectively. *Journal of Cell Communication and Signaling*, **12**(4), 625–629.

Pertea, M., Shumate, A., Pertea, G., … Salzberg, S. (2018). Thousands of large-scale RNA sequencing experiments yield a comprehensive new human gene list and reveal extensive transcriptional noise. *BioRxiv*, 332825.

Petit, F., Sears, K. E., & Ahituv, N. (2017). Limb development: A paradigm of gene regulation. *Nature Reviews Genetics*, pp. 245–258.

Petrenko, N., Jin, Y., Wong, K. H., & Struhl, K. (2016). Mediator Undergoes a Compositional Change during Transcriptional Activation. *Molecular Cell*, **64**(3), 443–454.

Plank, J. L., & Dean, A. (2014). Enhancer function: Mechanistic and genome-wide insights come together. *Molecular Cell*, pp. 5–14.

Pott, S., & Lieb, J. D. (2015). What are super-enhancers? *Nature Genetics*, pp. 8–12.

Poulet, B., Hamilton, R. W., Shefelbine, S., & Pitsillides, A. A. (2011). Characterizing a novel and adjustable noninvasive murine joint loading model. *Arthritis and Rheumatism*, **63**(1), 137–147.

Pradeepa, M. M., Grimes, G. R., Kumar, Y., … Bickmore, W. A. (2016). Histone H3 globular domain acetylation identifies a new class of enhancers. *Nature Genetics*, **48**(6), 681–686.

Proudfoot, N. (2016). Transcriptional termination in mammals: stopping the RNA polymerase II juggernaut. *Science*, **352**(6291), 11291-.

Rachfal, A. W., & Luquette, M. H. (2004). Expression of connective tissue growth factor (CCN2) in desmoplastic small round cell tumour. *J Clin Pathol*, **57**, 422–425.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., & Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, **470**(7333), 279–283.

Rahman, S., Zorca, C. E., Traboulsi, T., … Zenklusen, D. (2016). Single-cell profiling reveals that eRNA accumulation at enhancer-promoter loops is not required to sustain transcription. *Nucleic Acids Research*, **45**(6), 3017–3030.

Rao, S. S. P., Huang, S. C., Glenn St Hilaire, B., … Aiden, E. L. (2017). Cohesin Loss Eliminates All Loop Domains. *Cell*, **171**(2), 305–320.e24.

Rao, S. S. P., Huntley, M. H., Durand, N. C., … Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**(7), 1665–1680.

Reiter, F., Wienerroither, S., & Stark, A. (2017). Combinatorial function of transcription factors and cofactors. *Current Opinion in Genetics and Development*, pp. 73–81.

Ren, G., Jin, W., Cui, K., … Zhao, K. (2017). CTCF-Mediated Enhancer-Promoter Interaction Is a Critical Regulator of Cell-to-Cell Variation of Gene Expression. *Molecular Cell*, **67**(6), 1049–1058.e6.

Reynard, L. N., Ratnayake, M., Santibanez-Koref, M., & Loughlin, J. (2016). Functional characterization of the osteoarthritis susceptibility mapping to CHST11 - A bioinformatics and molecular study. *PLoS ONE*, **11**(7), e0159024.

Rhee, D. K., Marcelino, J., Baker, M., … Carpten, J. D. (2005). The secreted glycoprotein lubricin protects cartilage surfaces and inhibits synovial cell overgrowth. *Journal of Clinical Investigation*, **115**(3), 622–631.

Richmond, T. J., & Davey, C. A. (2003). The structure of DNA in the nucleosome core. *Nature*, **423**(6936), 145–150.

Rickels, R., Herz, H. M., Sze, C. C., … Shilatifard, A. (2017). Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. *Nature Genetics*, **49**(11), 1647–1653.

Rigueur, D., & Lyons, K. M. (2014). Whole-mount skeletal staining. *Methods in Molecular Biology*, **1130**, 113–121.

Roelofs, A. J., Zupan, J., Riemen, A. H. K., … De Bari, C. (2017). Joint morphogenetic cells in the adult mammalian synovium. *Nature Communications*, **8**

Roider, H. G., Kanhere, A., Manke, T., & Vingron, M. (2007). Predicting transcription factor affinities to DNA from a biophysical model. *Bioinformatics*, **23**(2), 134–141.

Rowley, M. J., Nichols, M. H., Lyu, X., … Corces, V. G. (2017). Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Molecular Cell*, **67**(5), 837–852.e7.

Roy, A. L., & Singer, D. S. (2015). Core promoters in transcription: Old problem, new insights. *Trends in Biochemical Sciences*, pp. 165–171.

Rubin, A. J., Barajas, B. C., Furlan-Magaril, M., … Khavari, P. A. (2017). Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nature Genetics*, **49**(10), 1522–1528.

Safadi, F. F., Xu, J., Smock, S. L., … Popoff, S. N. (2003). Expression of connective tissue growth factor in bone: Its role in osteoblast proliferation and differentiation in vitro and bone formation in vivo. *Journal of Cellular Physiology*, **196**(1), 51–62.

Salva, J. E., & Merrill, A. E. (2017, April 1). Signaling networks in joint development. *Developmental Dynamics*, John Wiley & Sons, Ltd, pp. 262–274.

Sawado, T., Halow, J., Bender, M. A., & Groudine, M. (2003). The β-globin locus control region (LCR) functions primarily by enhancing the transition from transcription initiation to elongation. *Genes and Development*, **17**(8), 1009–1018.

Sayre, M. H., Tschochner, H., & Kornberg, R. D. (1992). Reconstitution of transcription with five purified initiation factors and RNA polymerase II from Saccharomyces cerevisiae. *The Journal of Biological Chemistry*, **267**(32), 23376–82.

Schaukowitch, K., Joo, J. Y., Liu, X., Watts, J. K., Martinez, C., & Kim, T. K. (2014). Enhancer RNA facilitates NELF release from immediate early genes. *Molecular Cell*, **56**(1), 29–42.

Schwarzer, W., Abdennur, N., Goloborodko, A., … Spitz, F. (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, **551**(7678), 51–56.

Scruggs, B. S., Gilchrist, D. A., Nechaev, S., … Adelman, K. (2015). Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. *Molecular Cell*, **58**(6), 1101–1112.

Segarini, P. R., Nesbitt, J. E., Li, D., Hays, L. G., Yates, J. R., & Carmichael, D. F. (2001). The low density lipoprotein receptor-related protein/alpha2-macroglobulin receptor is a receptor for connective tissue growth factor. *The Journal of Biological Chemistry*, **276**(44), 40659–67.

Serra, R., Karaplis, A., & Sohn, P. (1999). Parathyroid hormone-related peptide (PTHrP)-dependent and -independent effects of transforming growth factor β (TGF-β)on endochondral bone formation. *Journal of Cell Biology*, **145**(4), 783–794.

Sheaffer, K. L., Kim, R., Aoki, R., … Kaestner, K. H. (2014). DNA methylation is required for the control of stem cell differentiation in the small intestine. *Genes and Development*, **28**(6), 652–664.

Shi-Wen, X., Howat, S. L., Renzoni, E. A., … Abraham, D. J. (2004). Endothelin-1 induces expression of matrix-associated genes in lung fibroblasts through MEK/ERK. *Journal of Biological Chemistry*, **279**(22), 23098–23103.

Shimo, T., Kanyama, M., Wu, C., … Koyama, E. (2004). Expression and roles of connective tissue growth factor in Meckel's cartilage development. In *Developmental Dynamics*, Vol. 231, John Wiley & Sons, Ltd, pp. 136–147.

Shimo, T., Kubota, S., Yoshioka, N., … Takigawa, M. (2006). Pathogenic role of connective tissue growth factor (CTGF/CCN2) in osteolytic metastasis of breast cancer. *Journal of Bone and Mineral Research*, **21**(7), 1045–1059.

Shimo, T., Nakanishi, T., Nishida, T., … Takigawa, M. (1999). Connective tissue growth factor induces the proliferation, migration, and tube formation of vascular endothelial cells in vitro, and angiogenesis in vivo. *Journal of Biochemistry*, **126**(1), 137–145.

Shlyueva, D., Stampfel, G., & Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nature Reviews. Genetics*, **15**(4), 272–86.

Shwartz, Y., Viukov, S., Krief, S., & Zelzer, E. (2016). Joint Development Involves a Continuous Influx of Gdf5-Positive Cells. *Cell Reports*, **15**(12), 2577–2587.

Sievers, F., Wilm, A., Dineen, D., … Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, **7**, 539.

Skuli, N., Liu, L., Runge, A., … Keith, B. (2009). Endothelial deletion of hypoxia-inducible factor-2α (HIF-2α) alters vascular function and tumor angiogenesis. *Blood*, **114**(2), 469–477.

Smerdel-Ramoya, A., Zanotti, S., Stadmeyer, L., Durant, D., & Canalis, E. (2008). Skeletal overexpression of connective tissue growth factor impairs bone formation and causes osteopenia. *Endocrinology*, **149**(9), 4374–4381.

Smith, E., & Shilatifard, A. (2014). Enhancer biology and enhanceropathies. *Nature Structural & Molecular Biology*, **21**(3), 210–9.

Soeda, T., Deng, J. M., De Crombrugghe, B., Behringer, R. R., Nakamura, T., & Akiyama, H. (2010). Sox9-expressing precursors are the cellular origin of the cruciate ligament of the knee joint and the limb tendons. *Genesis*, **48**(11), 635–644.

Song, J. J., Aswad, R., Kanaan, R. A., … Popoff, S. N. (2007). Connective tissue growth factor (CTGF) acts as a downstream mediator of TGF-β1 to induce mesenchymal cell condensation. *Journal of Cellular Physiology*, **210**(2), 398–410.

Sonnylal, S., Shi-wen, X., Leoni, P., … De Crombrugghe, B. (2010). Selective expression of connective tissue growth factor in fibroblasts in vivo promotes systemic tissue fibrosis. *Arthritis and Rheumatism*, **62**(5), 1523–1532.

Soutourina, J. (2017). Transcription regulation by the Mediator complex. *Nature Reviews Molecular Cell Biology*.

Spitz, F., & Furlong, E. E. M. (2012). Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics*, **13**(9), 613–626.

Spurrell, C. H., Dickel, D. E., & Visel, A. (2016). The Ties That Bind: Mapping the Dynamic Enhancer-Promoter Interactome. *Cell*, **167**(5), 1163–1166.

St-Jacques, B., Hammerschmidt, M., & McMahon, A. P. A. (1999). Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. *Genes & Development*, **13**(16), 2072–2086.

Stampfel, G., Kazmar, T., Frank, O., Wienerroither, S., Reiter, F., & Stark, A. (2015). Transcriptional regulators form diverse groups with context-dependent regulatory functions. *Nature*, **528**, 147–151.

Stasevich, T. J., Hayashi-Takanaka, Y., Sato, Y., … Kimura, H. (2014). Regulation of RNA polymerase II activation by histone acetylation in single living cells. *Nature*, **516**(7530), 272–275.

Storm, E. E., & Kingsley, D. M. (1996). Joint patterning defects caused by single and double mutations in members of the bone morphogenetic protein (BMP) family. *Development (Cambridge, England)*, **122**(12), 3969–79.

Struhl, K., & Segal, E. (2013). Determinants of nucleosome positioning. *Nature Structural and Molecular Biology*, pp. 267–273.

Taher, L., McGaughey, D. M., Maragh, S., … Ovcharenko, I. (2011). Genome-wide identification of conserved regulatory function in diverged sequences. *Genome Research*, **21**(7), 1139–49.

Tang, X., Muhammad, H., McLean, C., … Vincent, T. L. (2018). Connective tissue growth factor contributes to joint homeostasis and osteoarthritis severity by controlling the matrix sequestration and activation of latent TGFβ. *Annals of the Rheumatic Diseases*, **77**(9), 1372–1380.

Tang, Z., Luo, O. J., Li, X., … Ruan, Y. (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell*, **163**(7), 1611–1627.

The ENCODE Project Consortium. (2011). A User's Guide to the Encyclopedia of DNA Elements (ENCODE). *PLoS Biology*, **9**(4), e1001046.

The ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**(7414), 57–74.

Thomas-Chollier, M., Hufton, A., Heinig, M., … Vingron, M. (2011). Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs. *Nature Protocols*, **6**(12).

Thurman, R. E., Rynes, E., Humbert, R., … Stamatoyannopoulos, J. A. (2012). The accessible chromatin landscape of the human genome. *Nature*, **489**(7414), 75–82.

Tomita, N., Hattori, T., Itoh, S., … Takigawa, M. (2013). Cartilage-Specific Over-Expression of CCN Family Member 2/Connective Tissue Growth Factor (CCN2/CTGF) Stimulates Insulin-Like Growth Factor Expression and Bone Growth. *PLoS ONE*, **8**(3).

Tran, C. M., Shapiro, I. M., & Risbud, M. V. (2013). Molecular regulation of CCN2 in the intervertebral disc: Lessons learned from other connective tissues. *Matrix Biology*, **32**(6), 298–306.

Tuli, R., Tuli, S., Nandi, S., … Tuan, R. S. (2003). Transforming growth factor-beta-mediated chondrogenesis of human mesenchymal progenitor cells involves N-cadherin and mitogen-activated protein kinase and Wnt signaling cross-talk. *The Journal of Biological Chemistry*, **278**(42), 41227–36.

Van Beek, J. P., Kennedy, L., Rockel, J. S., Bernier, S. M., & Leask, A. (2006). The induction of CCN2 by TGFβ1 involves Ets-1. *Arthritis Research & Therapy*, **8**.

van den Bosch, M. H., Blom, A. B., van Lent, P. L., … van den Berg, W. B. (2014). Canonical Wnt signaling skews TGF-β signaling in chondrocytes towards signaling via ALK1 and Smad 1/5/8. *Cellular Signalling*, **26**, 951–958.

van der Kraan, P. M., Blaney Davidson, E. N., Blom, A., & van den Berg, W. B. (2009). TGF-beta signaling in chondrocyte terminal differentiation and osteoarthritis. Modulation and integration of signaling pathways through receptor-Smads. *Osteoarthritis and Cartilage*, pp. 1539–1545.

Van der Kraan, P. M., & Van den Berg, W. B. (2012). Chondrocyte hypertrophy and osteoarthritis: Role in initiation and progression of cartilage degeneration? *Osteoarthritis and Cartilage*, **20**(3), 223–232.

Visel, A., Blow, M. J., Li, Z., … Pennacchio, L. A. (2009a). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, **457**(7231), 854–858.

Visel, A., Bristow, J., & Pennacchio, L. A. (2007a). Enhancer identification through comparative genomics. *Seminars in Cell and Developmental Biology*, pp. 140–152.

Visel, A., Minovitsky, S., Dubchak, I., & Pennacchio, L. A. (2007b). VISTA Enhancer Browser - A database of tissue-specific human enhancers. *Nucleic Acids Research*, **35**

Visel, A., Rubin, E. M., & Pennacchio, L. A. (2009b). Genomic views of distant-acting enhancers. *Nature*, **461**(7261), 199–205.

Vortkamp, A., Lee, K., Lanske, B., Segre, G. V, Kronenberg, H. M., & Tabin, C. J. (1996). Regulation of rate of cartilage differentiation by Indian Hedgehog and PTH-related protein.

Science, American Association for the Advancement of Science, pp. 613–622.

Wang, W., Rigueur, D., & Lyons, K. M. (2014). TGFβ signaling in cartilage development and maintenance. *Birth Defects Research Part C - Embryo Today: Reviews*, **102**(1), 37–51.

Wang, Y., Song, F., Zhang, B., … Yue, F. (2018). The 3D Genome Browser: A web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biology*, **19**(1).

Wei, G., Srinivasan, R., Cantemir-Stone, C. Z., … Ostrowski, M. C. (2009). Ets1 and Ets2 are required for endothelial cell survival during embryonic angiogenesis. *Blood*, **114**(5), 1123–1130.

Weintraub, A. S., Li, C. H., Zamudio, A. V, … Young, R. A. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell*, **171**(7), 1573–1588.e28.

Wellik, D. M., & Capecchi, M. R. (2003). Hox10 and Hox11 genes are required to globally pattern the mammalian skeleton. *Science*, **301**(5631), 363–367.

Whalen, S., Truty Rebecca M, & Pollard, K. S. (2016). Enhancer–promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nature Genetics*, **48**(5), 488–497.

Whyte, W. A., Orlando, D. A., Hnisz, D., … Young, R. A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, **153**(2), 307–319.

Wilusz, R. E., Sanchez-Adams, J., & Guilak, F. (2014). The structure and function of the pericellular matrix of articular cartilage. *Matrix Biology*, pp. 25–32.

Wright, E., Hargrave, M. R., Christiansen, J., … Koopman, P. (1995). The Sry-related gene Sox9 is expressed during chondrogenesis in mouse embryos. *Nature Genetics*, **9**(1), 15–20.

Wu, M., Chen, G., & Li, Y. P. (2016). TGF-β and BMP signaling in osteoblast, skeletal development, and bone formation, homeostasis and disease. *Bone Research*, **4**.

Xie, D., Nakachi, K., Wang, H., Elashoff, R., & Koeffler, H. (2001). Elevated levels of connective tissue growth factor, WISP-1 and CRY61 in primary breast cancers associated with more advanced features. *Cancer Research*, **61**, 8917–8923.

Xu, H., Xiao, T., Chen, C.-H., … Liu, X. S. (2015). Sequence determinants of improved CRISPR sgRNA design. *Genome Research*, **25**(8), 1147–57.

Yamaai, T., Nakanishi, T., Asano, M., … Takigawa, M. (2005). Gene expression of connective tissue growth factor (CTGF/CCN2) in calcifying tissues of normal and cbfa1-null mutant mice in late stage of embryonic development. *Journal of Bone and Mineral Metabolism*, **23**(4), 280–288.

Yan, J., Chen, S. A. A., Local, A., … Ren, B. (2018). Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers. *Cell Research*, **28**(2), 204–220.
Yang, D., Jang, I., Choi, J., … Lee, B. (2018). 3DIV: A 3D-genome Interaction Viewer and database. *Nucleic Acids Research*, **46**(2).7

Yang, L., Tsang, K. Y., Tang, H. C., Chan, D., & Cheah, K. S. E. (2014). Hypertrophic chondrocytes can become osteoblasts and osteocytes in endochondral bone formation. *Proceedings of the National Academy of Sciences*, **111**(33), 12097–12102.

Yao, B., Wang, Q., Liu, C.-F., … Lefebvre, V. (2015). The SOX9 upstream region prone to chromosomal aberrations causing campomelic dysplasia contains multiple cartilage enhancers. *Nucleic Acids Research*, **43**(11), 5394–5408.

Yao, Y., & Wang, Y. (2013). ATDC5: An excellent in vitro model cell line for skeletal development. *Journal of Cellular Biochemistry*, **114**(6), 1223–1229.

Yeger, H., & Perbal, B. (2007). The ccn family of genes: A perspective on ccn biology and therapeutic potential. *Journal of Cell Communication and Signaling*, **1**(3–4), 159–164.

Yi, S. E., Daluiski, A., Pederson, R., Rosen, V., & Lyons, K. M. (2015). The type i BMP receptor BMPRIB is required for chondrogenesis in the mouse limb. *Development*, **127**, 621–630.

Yin, J. -w., & Wang, G. (2014). The Mediator complex: a master coordinator of transcription and cell lineage development. *Development*, **141**(5), 977–987.

Yoshida, C. A., Yamamoto, H., Fujita, T., … Komori, T. (2004). Runx2 and Runx3 are essential for chondrocyte maturation, and Runx2 regulates limb growth through induction of Indian hedgehog. *Genes & Development*, **18**(8), 952–963.

Zabidi, M. A., Arnold, C. D., Schernhuber, K., … Stark, A. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature*, **518**(7540), 556–559.

Zabidi, M. A., & Stark, A. (2016). Regulatory Enhancer–Core-Promoter Communication via Transcription Factors and Cofactors. *Trends in Genetics*, pp. 801–814.

Zaucke, F., Dinser, R., Maurer, P., & Paulsson, M. (2001). Cartilage oligomeric matrix protein (COMP) and collagen IX are sensitive markers for the differentiation state of articular primary chondrocytes. *Biochemical Journal*, **358**(1), 17.

Zhou, X., von der Mark, K., Henry, S. S., … Bellizzi, J. (2014). Chondrocytes Transdifferentiate into Osteoblasts in Endochondral Bone during Development, Postnatal Growth and Fracture Healing in Mice. *PLoS Genetics*, **10**(12), e1004820.

Zhu, H., & Bendall, A. J. (2009). Dlx5 Is a Cell Autonomous Regulator of Chondrocyte Hypertrophy in Mice and Functionally Substitutes for Dlx6 during Endochondral Ossification. *PLoS ONE*, **4**(11), 8097.

Zhu, J., Adli, M., Zou, J. Y., … Bernstein, B. E. (2013a). Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell*, **152**(3), 642–654.

Zhu, Y., Sun, L., Chen, Z., Whitaker, J. W., Wang, T., & Wang, W. (2013b). Predicting enhancer transcription and activity from chromatin modifications. *Nucleic Acids Research*, **41**(22), 10032–10043.